

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 W O R L D  
 \*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mprch\_dp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Sun Sep 3 12:53:12 2000; MasPar time 29.26 Seconds  
 Tabular output not generated. 900.995 Million cell updates/sec

Title: >US-09-332-522B-8  
 Description: (1.1113) from US09332522B.pep  
 Perfect Score: 8012  
 Sequence: 1 MDLTLMNLIDAPLDESMDLF.....LKECYOLMKSLGNGISVKA 1113

Scoring table: PAM 150  
 Gap 11

Searched: 188963 seqs, 23586105 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 1:geneseqp

Statistics: Mean 40.405; Variance 208.262; scale 0.194

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1494	18.5	1147	1	Human SREBP-1a.	4.82e-108
2	1328	16.6	1141	1	R66390	2.09e-94
3	186	2.3	618	1	Human chromosome X TFE	6.04e-04
4	172	2.1	597	1	Human PRCC-TFE3 con	5.44e-03
5	172	2.1	700	1	Human NonO/TFE3 fusion	5.44e-03
6	172	2.1	834	1	Human PRCC-TFE3 fusion	5.44e-03
7	153	2.0	383	1	Petunia Ph6 gene produ	2.19e-02
8	149	1.9	160	1	Max 11.	1.83e-01
9	149	1.9	2414	1	Transcription factor p	1.83e-01
10	149	1.9	2414	1	Cellular transcription	1.83e-01
11	148	1.8	151	1	Max 14.	2.13e-01
12	140	1.7	160	1	Breast cancer associat	1.45e+00
13	135	1.7	258	1	DEL protein.	1.68e+00
14	134	1.7	644	1	A tumour suppressor ge	1.94e+00
15	133	1.7	1004	1	Human polyhomocytic 1 (	1.94e+00
16	133	1.7	1004	1	Elmeria cell surface a	6.13e+00
17	126	1.6	259	1	Cotton fibrous tissue	3.46e+00
18	125	1.6	297	1	Mycobacterium tubercu	3.46e+00
19	129	1.6	572	1	Human CLOCK protein.	5.32e+00
20	129	1.6	763	1	Human HSCLOCK polyep	5.32e+00
21	126	1.6	846	1	Mouse CLOCK protein.	5.32e+00
22	126	1.6	846	1		
23	126	1.6	855	1		

## ALIGNMENTS

RESULT ID	1	ALIGNMENTS
AC	R66390: standard; Protein; 1147 AA.	
DT	04-AUG-1995 (first entry)	
DE	Human SREBP-1a.	
KW	Probe: sterol regulatory element binding protein; regulatory protein;	
KW	cholesterol metabolism; sterol regulatory element-1; SREBP: SRE-1;	
KW	SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasma;	
KW	transcription factor; low density lipoprotein; LDL; receptor;	
KW	3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;	
KW	hypercholesterolaemia; enhancer.	
OS	Homo sapiens.	
FH	Key	
FT	Region	Location/Qualifiers
FT	FT	1..60
FT	FT	/note= "Acidic region"
FT	FT	61..177
FT	FT	/note= "Pro/Ser rich domain"
FT	FT	324..394
FT	FT	/label= bHLH-2ip
FT	FT	324..333
FT	FT	/note= "Basic region"
FT	FT	337..349
FT	FT	/note= "Helix 1"
FT	FT	358..373
FT	FT	/note= "Helix 2"
FT	FT	359..401
FT	FT	/note= "Putative Leucine zipper"
FT	FT	427..461
FT	FT	/note= "Ser/Gly/Pro rich domain"
FT	FT	462..1147
FT	FT	/note= "C-terminal domain"
PN	W09426922-A.	
PD	24-NOV-1994.	
PR	13-MAY-1994: UC5300.	
PR	01-OCT-1993: US-131365.	
PA	(TEXA ) UNIV TEXAS SYSTEM.	
PI	Briggs MR, Brown MS, Goldstein JL, Wang X;	
DR	WPL: 95-006813/01.	
DR	N-PSDB: 079037.	
PT	New sterol regulator element binding protein - used to develop	
PT	prods. and screening assays for agents for reducing plasma	
PT	cholesterol levels (Eng)	
PS	Claim 8; Page 185-98; 305pp; English.	
CC	The sequences given in R66390-91 represent the sterol regulatory element	
CC	binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory	
CC	proteins which are involved in the regulation of genes involved in	

24	122	1.5	166	1	R35450	Human eps8.	9.37e+00
25	121	1.5	234	1	R66913	Cotton fiber-specific	1.08e+01
26	120	1.5	212	1	W01019	Apoptosis-blocking pro	1.24e+01
27	120	1.5	232	1	W94346	Human Bcl-2 mutant pro	1.24e+01
28	119	1.5	233	1	W14590	Streptococcus pneumoni	1.43e+01
29	119	1.5	346	1	W60569	Human chromosome 19 de	1.43e+01
30	121	1.5	394	1	P70434	Amino acids 71-464 of	1.08e+01
31	121	1.5	394	1	R04834	N-myc protein	1.08e+01
32	122	1.5	412	1	W90060	Human basic helix-loop	9.37e+00
33	119	1.5	458	1	W14592	Streptococcus pneumoni	1.43e+01
34	124	1.5	546	1	W37888	Tobacco PABF protein.	7.06e+00
35	120	1.5	559	1	W89803	Staphylococcus aureus	1.24e+01
36	120	1.5	610	1	R10923	Maize protein encoded	1.43e+01
37	119	1.5	743	1	R51110	Human TLE-2.	1.43e+01
38	119	1.5	743	1	R51477	Human TLE-2.	1.43e+01
39	118	1.5	907	1	R80144	EBV gp350/220.	1.64e+01
40	120	1.5	940	1	R07070	Fibronectin-binding pr	1.24e+01
41	119	1.5	2441	1	W40058	Cellular transcription	1.43e+01
42	118	1.5	2441	1	R79054	CREB binding protein.	1.43e+01
43	119	1.5	2482	1	W23996	Human mitotin amino ac	1.43e+01
44	119	1.5	2482	1	R72826	Human mitotin.	1.43e+01
45	119	1.5	3248	1	R99795	Kinetochoe protein CE	1.43e+01

CC cholesterol metabolism that are under the control of an associated  
 CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins  
 CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members  
 CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-zip)  
 CC transcription factors. Each have the ability to bind to SRE sequences  
 CC and modulate SRE-mediated transcription. SRE-1 is a conditional enhancer  
 CC found in the promoters for the low density lipoprotein (LDL) receptor  
 CC and 3-hydroxy-3-methylglutaryl coenzyme A synthase genes. It increases  
 CC transcription in the absence of sterols and is inactivated when sterols  
 CC accumulate. Human SREBP-2 contains 1141 amino acids and has 47% identity  
 CC with SREBP-1a. SREBP-1a was the first recognised member of this family  
 CC and has 1147 amino acids. The resemblance between SREBP-1a and SREBP-2  
 CC includes an acidic N-terminus, a highly conserved bHLH-zip motif (71%  
 CC identical), and an unusually long extension of 740 amino acids on the  
 CC C-terminal side of the bHLH-zip region. SREBP-2 possesses one feature  
 CC lacking in SREBP-1a, a Glu rich region (27% Glu over 121 residues).  
 CC SREBP promotes SRE-1-mediated gene transcription, eg. LDL receptor  
 CC production in the presence of sterols. SREBP identified in screening  
 CC assays, may be used to reduce plasma cholesterol levels and in  
 CC controlling hypercholesterolaemia and its associated diseases.  
 CC Sequence 1147 AA;

Query Match 18.6%; Score 1494; DB 1; Length 1147;

Best Local Similarity 33.6%; Pred. No. 4,82e-108; Mismatches 388; Conservative 227; Indels 55; Gaps 51;

Db 169 PGFSTGSPGNTQOPLPGLPLASPPGVPVSLHTQ-V-QSVVPOQLLTVAAPTAAPVT 226  
 Oy 126 PTGLKAQPTATIHMDQRMPTNAVYPSLGSFVYQSMSPRTS-PVESANQNVNVM 184  
 Db 227 TTVTSQIQOVPVLDPHFKADSLILTAKTGTGATYKAGSLPYSVGTTCVQCPRLTVLS 286  
 Oy 185 QPVAATPAPASAPRLPOQSY-PQPFITYNSKA-GMTSDEA-MYLLQPTVAPSPTPSPVAP 241  
 Db 287 GGTITATVPLVYDAEKL-PI-NRLAAGSKAPAS-AQSRG-E-KRTAHNAIEKRYSSIND 341  
 Oy 242 PPTSTGSRASKVAVAPLAPSPAAEMVQGVPIINRVQPKVEKRSKHNHAIERYRTSIND 301  
 Db 342 KIIEKDLVVGTEAKLNKSAVLRAKADYIRFLQHSNOKLQENTSL-RTAVHKSRS-LMD 399  
 Oy 302 KINELKNLVYGQAKLNKSAVLKRSIDKTRDQROHDKAELOKRELMARSDSKYVD 361  
 Db 400 LVS-ACGSGGNDVLMGKYKT-EVEDTLTPPPSDAGSPQSSPPLSGSGSGSGSGSDS 457  
 Oy 362 LIQGTTRPGRAKKRRESQGTFTTAGLTPRPSDESPSL-S-PMH-SDLSLPPSPYGGST 419  
 Db 458 EPDSPVFEEDSKAKPDRQPLHSRGLDRSLALCTVFLCTCNPLASLGAARLPSPSD 517  
 Oy 420 ASCSSGSSSSNEEPLVPS-SMRGMAHTRSLGICMFMFAILAVNPFKTYL-QRGHYDSND 477  
 Db 518 TTSVYHSGRNVLGTESRDGPMAQMLPPVYVNLGLLVLSVLYLVFYGEPVTRPHSG 577  
 Oy 478 DLGDMGQ-RRLTSYDV-BEGEPFAVWQSSWIMLNFIMLGICVLYLVYGPQDADQND 535  
 Db 578 PAVYFWRHKKQDLDLAKDFPQAAQQLWLA-LRALGRPLPTSHDLACSLL-WNLIN 634  
 Oy 536 -A-YC-OHRORADFFYSQSSQAYAG-YLNCILHMGSLPLASRLE-CYLQTTQWFLNF 589  
 Db 635 LIORLWGVWMLAGRAGGLQOQCALRYDASARADALVYHKLHOLHTMKKHGHHTAN 694  
 Oy 590 LFRHMLGVLSRRSGGLPSNAASRQALASARELALLNRLNQLQLTGSGRGDMNGIM 649  
 Db 695 LALSALNTAEACGDAVSVATLAEIYVAALRYKTSLPRALH-FLTFEFLSARQAC-LAQ 752  
 Oy 650 MALFSSNAEVAHNLITRETICIHVMTALRMKRSAPKMLQDFPARYYSRAREGGRIR 709  
 Db 733 SGSVPAQWICLPVGHREFDVGDMSV-LSTP-WESIVSLAGNPVPLAQTOLFRHL 809  
 Oy 710 ATEQOELRMATFAYGYRCATHVFTYDSDGEGDGFTRLRNCPDPAHYTIKQYREHL 769  
 Db 810 LERALNCTQPPSPGSGADGKEFSPALGYDOLINSCSMAAGAPAYSFSSSM-ATTGG 868  
 Oy 770 LFKSIQCLVAGAKHSGGLPTSSVSGEAEOLQOQOHSGITVSNVLYKTSILKDTLWADDE 829

Db 869 VDPVAKMASITRAVYIHWLRDEAEERLCPVLEHLPVIOE-SER-PLPRAAHS-FKA 925  
 Oy 830 RDTNVVWMADEVLETAVHLLGEDTLAEOLYGRIKOMPTOLOQCGENDHLPKA-LHAVALRA 888  
 Db 926 ARALL-GGAKA-ESGPALT-I-CEKAGYIQDSLATTPASSSID-KAV-QLFLCDLLV 979  
 Oy 889 KMILKNNMADLSKLOKVLNDESSVELOECTVIRITDAGIKILFOLLTCDWLE 948  
 Db 960 VTSLSMROQPPAPAPAAQGSRRQASALEGRFORLSSLRRLAQSFRAMRRVFLHE 1039  
 Oy 949 TITALM-ELEHNN-M-EDDGYQVP-GEVLE-K-FQTLNLSLRNVENIPNAGSYIYE 1002  
 Db 1040 ATARLMAGASPTTHOLLDSRLRRRAGPGKG-GAVAELEPR-PTREHAEALILASYC 1096  
 Oy 1003 AVCRLMAGASCPITQOLLDRSLSRSHASIFCGSKDRQONFVGEGERERASAMYACKY 1062  
 Db 1097 LPPGLTAPGQGVGLAFAARTLEKGRRLIHDQOQMLMGG 1141  
 Oy 1063 LPPALLSSPGERGMLAEAAKTLERVGDKRKLECYOLMKSLGNG 1107

RESULT 2

ID R66391 standard; Protein; 1141 AA.

AC R66391.  
 DT 04-AUG-1995 (first entry)  
 DE Human SREBP-2.  
 KW Probe: sterol regulatory element binding protein; regulatory protein;  
 KW cholesterol metabolism; sterol regulatory element-1; SREBP; SRE-1;  
 KW SREBP-1; SREBP-2; basic-helix-loop-helix-leucine zipper; plasm;  
 KW transcription factor; low density lipoprotein; LDL; receptor;  
 KW 3-hydroxy-3-methylglutaryl coenzyme A synthase; cholesterol;  
 KW hypercholesterolaemia; enhancer.  
 OS Homo sapiens.  
 FH Key  
 FH region  
 FT 1. 50  
 FT /note- "Acidic region"  
 FT domain  
 FT 51. 124  
 FT /note- "Ser/gly/Pro rich domain"  
 FT domain  
 FT 125. 246  
 FT /note- "Gln rich domain"  
 FT region  
 FT 331. 401  
 FT /label- bHLH-zip  
 FT region  
 FT 331. 340  
 FT /note- "Basic region"  
 FT region  
 FT 344. 356  
 FT /note- "Helix 1"  
 FT region  
 FT 363. 381  
 FT /note- "Helix 2"  
 FT region  
 FT 382. 401  
 FT /note- "Putative leucine zipper"  
 FT domain  
 FT 458. 1141  
 FT /note- "C-terminal domain"  
 PN W09426922-A.  
 PD 24-NOV-1994.  
 PF 13-MAY-1994; U05300.  
 PR 13-MAY-1993; US-061967.  
 PR 01-OCT-1993; US-131365.  
 PA (TEXTA) UNIV TEXAS SYSTEM.  
 PI Briggs MR, Brown MS, Goldstein JL, Wang X;  
 DR WPI; 95-006813/01.  
 DR N-PSDB; Q79038.  
 DT New sterol regulator element binding protein - used to develop  
 PT prods. and screening assays for agents for reducing plasma  
 PT cholesterol levels (Eng)  
 PS Claim 10; Page 227-36; 305pp; English.  
 CC The sequences given in R66390-91 represent the sterol regulatory element  
 CC binding proteins (SREBP), SREBP-1a and SREBP-2. SREBP's are regulatory  
 CC proteins which are involved in the regulation of genes involved in  
 CC cholesterol metabolism that are under the control of an associated  
 CC sterol regulatory element-1 (SRE-1) enhancer sequence. SREBP proteins  
 CC fall into two families, SREBP-1 and SREBP-2. Both proteins are members  
 CC of a family of basic-helix-loop-helix-leucine zipper (bHLH-zip)





Dd	349	DNNLEERRRRRINDRIKETLTLPKSSDDEMRNKKITLKAASYDIRKLQEOGRS - 407
Oy	286	SAHNAIEREYRSINDKINELKNLY--VGE-QAKLNKSAVLKRSDIKTRDLQRONHDLA 342
Dd	408	DLESQRSL 416
Oy	343	ELORLQREL 351
RESULT	4	
ID	W52821	standard; Protein: 597 AA.
AC	M52821;	
DT	08-SEP-1998	(first entry)
DE	Human PRCC-TFE3 construct protein from cell lines UOK120 and UOK146.	
KM	PRCC; papillary renal cell carcinoma; TFE3; transcription factor;	
XW	fusion protein; translocation; diagnosis; treatment.	
OS	Homo sapiens.	
FH	Synthetic.	
FT	Key	Location/Qualifiers
FT	Protein	1..597
FT		/label= "PRCC-TFE3"
FT		/note= "Fusion protein"
FT	Misc_difference	1..156
FT		/label= PRCC
FT		*
FT	Misc_difference	159..597
FT		/note= "Papilloma renal cell virus partial sequence"
FT		/label= TFE3
FT		/note= "transcription factor partial sequence"
FN	WO9806871-A1.	
PD	19-FEB-1998.	
PF	13-AUG-1997; G02209.	
PA	13-AUG-1996; GB-016986.	
PI	(CANC-) CANCER RES CAMPAIGN TECHNOLOGY.	
DR	Clark J, Cooper C, Shipley J;	
N-PDB:	WPI: 98-139537/14.	
PT	N-PDB: V20957.	
PT	Diagnosing papillary renal cell carcinoma by detecting gene	
PT	trans-location - resulting in fusion of TFE3 gene with some other	
PT	gene, also related vectors, transformed cells' specific binding	
PT	reagents, peptide(s) encoded by fusions and therapeutic anti-sense	
PS	sequences	
PS	Claim 9; Fig 2B; 71pp; English.	
CC	This sequence represents a novel fusion protein constructed from a	
CC	papillary renal cell carcinoma (PRCC) associated protein and the	
CC	transcription factor TFE3 which is found in cell lines UOK120 and UOK146	
CC	and used in a method for the diagnosis, prophylactic and therapeutic	
CC	treatment of papillary renal cell carcinoma. The translocation t(x);1	
CC	(p11.2;q21.2) found in PRCC results in a fusion of the TFE3 gene with a	
CC	new chromosome 1 gene designated PRCC (at 1q21.2), resulting in	
CC	expression of a fusion protein between the N-terminus of PRCC and almost	
CC	the whole of the TFE3 gene. Normal TFE3 transcripts are no longer	
CC	produced. Two other fusion partners for TFE3 have also been detected;	
CC	NonO, from a invx (p11.2; q13-24 or 12) translocation and the PSF splice	
CC	factor gene, resulting in t(x);1 (p11.2;p34). These trans-locations	
CC	define a subgroup of PRCC generally encountered in patients younger	
CC	than 25.	
SO	Sequence 597 AA:	
Query Match	2.18;	Score 172; DB 1; Length 597;
Best Local Similarity	40.8%;	Pred. No. 5,44e+03;
Matches 29; Conservative	19;	Mismatches 19; Indels 4; Gaps 3;
Dd	326	KRDNNLEIRRRRNNDRIKEGLTIPKSSPEEMRWKGTILKAASYDIRKLQEOGRS 385
Oy	284	KRAHNAIERRTYSINDKINELKNLY--VGE-QAKLNKSAVLKRSDIKTRDLQRONHDL 340
Dd	386	K-DLESQRSL 395
Oy	341	KAELORLQREL 351
RESULT	5	

ID	W52822	standard; Protein; 700 AA
DE	08-SEP-1998	(first entry)
DE	Human Nono/TFE3 fusion product from cell line UOK109.	
KW	PRCC; papillary renal cell carcinoma; TFE3; transcription factor;	
KW	fusion protein; translocation; diagnosis; treatment; Nono; p54-nrb.	
OS	Homo sapiens.	
OS	Synthetic.	
FT	Key	
FT	Protein	
FT	1..700	
FT	/label= "Nono/TFE3"	
FT	/note= "Fusion protein"	
FT	Misc.difference 1..377	
FT	/label= Nono	
FT	/note= "also known as p54-nrb"	
FT	Misc.difference 378..700	
FT	/label= TFE3	
FT	/note= "transcription factor sequence"	
PN	M09806871-A1.	
PD	19-FEB-1998.	
PF	13-AUG-1997; G02209.	
PR	13-AUG-1996; GB-016986.	
PA	(CANC- ) CANCER RES CAMPAIGN TECHNOLOGY.	
PI	Clark J, Cooper C, Shipley J;	
DR	WPI: 98-159557/14.	
DR	N-PSDB: V20958	
PT	Diagnosing papillary renal cell carcinoma by detecting gene	
PT	trans-location - resulting in fusion of TFE3 gene with some other	
PT	gene, also related vectors, transformed cells, specific binding	
PT	reagents, peptide(s) encoded by fusions and therapeutic anti-sense	
PT	sequences	
PS	Claim 9; Fig 3A; 71pp; English.	
CC	This sequence represents a novel fusion protein constructed from the Nono	
CC	protein (also known as p54-nrb) and the transcription factor TFE3 which	
CC	is found in cell line UOK109 and used in a method for the diagnosis,	
CC	prophylactic and therapeutic treatment of papillary renal cell carcinoma.	
CC	The translocation t(X;1) (p11.2;q21.2) found in papillary renal cell	
CC	carcinoma (PRCC) associated protein (PRCC) results in a fusion of the	
CC	TFE3 gene with a new chromosome 1 gene designated PRCC (at 1q21.2),	
CC	resulting in expression of a fusion protein between the N-terminus of	
CC	PRCC and almost the whole of the TFE3 gene. Normal TFE3 transcripts are	
CC	no longer produced. Two other fusion partners for TFE3 have also been	
CC	detected: Nono..from a invx (p11.2; q13-24 or 12) translocation and the	
CC	PSF splice factor gene, resulting in t(X;1) (p11.2;p34). These	
CC	trans-locations define a subgroup of PRCC generally encountered in	
CC	patients younger than 25.	
SO	Sequence 700 AA;	
Query Match	2.18; Score 172; DB 1; Length 700;	
Best Local Similarity	40.68; Pred. No.5,44e-03;	
Matches 29; Conservative	19; Mismatches 19; Indels 4; Gaps 3;	
Db	429 KKHNNHIERRRRNINDRIKELGTLPKSPDEKRMNGRTLKASVDYIRKLQEOORS 488	
Oy	284 KRSHNNHIERRRYTSINDKINELKNV--VGE-QAKLNKSAVLRKSIDKINDLORONHDL 340	
Db	489 K-DLESQRSL 498	
Oy	341 KAEIQRLOREL 351	
RESULT	6	
ID	W52820	
AC	W52820;	
DT	08-SEP-1998 (first entry)	
DE	Human PRCC-TFE3 fusion protein from cell line UOK124.	
KW	PRCC; papillary renal cell carcinoma; TFE3; transcription factor;	
KW	fusion protein; translocation; diagnosis; treatment.	
OS	Homo sapiens.	
OS	Synthetic.	
FT	Key	
FT	Protein	
FT	1..834	
FT	/label= PRCC-TFE3	
FT	Location/Qualifiers	
FT	1..834	
FT	/label= PRCC-TFE3	

```

FT      Misc_ difference 1.393 /note= "fusion protein"
FT      /label= PRCO
FT      /note= "Papillary renal cell carcinoma partial
FT      Misc_difference 394.834 /label= TFE3
FT      /note= "Transcription factor partial sequence"
PN      W09806871-A1.
PD      19-FEB-1998.
PR      13-AUG-1997: G02209.
PR      13-AUG-1996: GB-016986.
PA      (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI      Clark J, Cooper C, Shipley J:
DR      WPJ: 98-159557/14.
DR      N-PDSB: V20956.
PT      Diagnosing papillary renal cell carcinoma by detecting gene
PT      translocation - resulting in fusion of TFE3 gene with some other
PT      gene, also related vectors, transformed cells, specific binding
PT      reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT      sequences
PS      Claim 9: Fig 2A: 71pp: English.
CC      This sequence represents a novel fusion protein constructed from a
CC      papillary renal cell carcinoma (PRCO) associated protein and the
CC      transcription factor TFE3 which is found in cell line UOK124 and used in
CC      a method for the diagnosis, prophylactic and therapeutic treatment of
CC      papillary renal cell carcinoma. The translocation t(X;1)(p11.2;q21.2)
CC      found in PRCO results in a fusion of the TFE3 gene with a new chromosome
CC      1 gene designated PRCO (at 1q21.2), resulting in expression of a fusion
CC      protein between the N-terminus of PRCO and almost the whole of the TFE3
CC      gene. Normal TFE3 transcripts are no longer produced. Two other fusion
CC      partners for TFE3 have also been detected: Nono, from a InvX (p11.2;
CC      q13-24 or 12) translocation and the PSF splice factor gene, resulting
CC      in t(X;1)(p11.2;p24). These translocations define a subgroup of PRCO
CC      generally encountered in patients younger than 25.
SQ      Sequence 834 AA:

Query Match          2.1%; Score 172; DB 1; Length 834;
Best Local Similarity 40.8%; Pred. No. 5.44e-03;
Matches 29; Conservative 19; Mismatches 19; Indels 4; Gaps 3;

Db      563 KKDHNHIERRRNRINDRIKELGTLPKSSDPEMNKGITIIKASVDYIRIKLOEQORS 622
       1 : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      284 KRSHNHNIEERRYTRISINDKIINLVNV--VGE-QAKLNKSAVLKSIDIKRIDLORONHD L 340
Oy      341 KAELQRIQLREL 351

Db      623 K-DLESRORSL 632
       1 : | : | | |
Oy      341 KAEIQRILREL 351

RESULT 7
ID      R62656 standard: Protein; 383 AA.
AC      R62656:
DT      08-JUN-1995 (first entry)
KW      Petunia ph6 gene product.
KW      Vacuolar pH: pH gene; Petunia.
OS      Petunia hybrida strain Y26.
FH      Key location//Nullifiers
       192..198
FT      /label= Helix-loop-helix structural motif
FT      /note= "see also Z39..245"
FN      W09423561-A.
PD      27-OCT-1994.
PD      15-APR-1994: U04173.
PR      16-APR-1993: US-049282.
PA      (DNAP ) DNA PLANT TECHNOLOGY CORP.
PI      Chujak GS, Courtney-guterson N, Dooner HK, Keller J:
PI      Nijjar CS, Ralston EJ:
DR      WPJ: 94-341349/42.
DR      N-PDSB: 073000.
PT      vacuolar pH gene and constructs containing it - for alteration of
PT      vacuolar pH used in the formation of blue flowers
OS      Disclosure: Page 42-44: 62pp: English.
```

	CC	The V26 fibrin of Petunia was used. Poly A RNA was isolated from
	CC	total RNA from flower buds and used to generate a cDNA library in
	CC	the vector lambda ZapII (Stratagene). The SstI to BamHI fragment at
	CC	the left hand side of Ac was used to isolate petA1-1 (contg. pH6
	CC	cDNA), the sequence for which is claimed (see 073000 FT). The AA
	CC	sequence comprises a helix-loop-helix structural motif starting with
	CC	the sequence NIVLAER (starting at residue 192) and extending to the
	CC	sequence KKKVQDLE (ending at residue 245). Proteins including this
	CC	motif include the myc family of oncogenes, regulators of neuron and
	CC	muscle development, and regulators of segmentation and organ
	CC	patterning in Drosophila. In plants, the motif is found in R(S).
	CC	a protein involved with regulating anthocyanin synthesis in Maize.
	CC	(Comparison of the pH6 and R(S) sequences over the 54 AA helix-loop-
	CC	helix region detected 57% AA identity. Outside this region only 1%
	CC	identity was found.
	SQ	Sequence 383 AA:
	Query Match	2.0% Score 163; DB 1; Length 383;
	Best Local Similarity 21.1%; Pred. No. 2,19e+02;	
	Matches 30; Conservative 45; Mismatches 66; Indels 1; Gaps 1;	
Dd	160	VSPKSRDATVDSSASTAFRCGSIQDEPFGNHYLAERRRRKLNERFIILSLYFV- 218   :   :::  :
Qy	255	VAPLAPAPAMEVGCKPINRVQPKVEVKRSHNMIERIRPSINDKIETLNKLVGEQ 314   :::  :
Dd	219	TKMDKASILGDITYYKOLRKKYODLEFARNQTEATLTQTGTCTVYKLOGRGKRMMKIVE 278   :::  :
Qy	315	AKLKRSAYVLKRSIDIKINDLORONHDLKAELQRLQRELMADSGSKVXDLLQTRPCRAK 374   :::  :
Dd	279	GSVGGGCAKITASSPSTTHEE 300   :::  :
Qy	375	KRRSSQTFPTTDAGLRPPRSD 396   :::  :
RESULT	8	
ID	W10041	standard; Protein; 160 AA.
AC	W10041;	
DT	07-FEB-1998	(first entry)
DE	Max 11.	
KM	muringe; mSina; mammalian homologue; Saccharomyces cerevisiae; repressor;	
KW	Sln3; Mad; Max; msln:Mad complex; mSln:Mad:Max complex; MYC; promoter;	
KW	basic helix-loop-helix zipper protein; compete; DNA-binding;	
KW	MYC:Max complex; activate; transcription; gene regulation.	
OS	Homo sapiens.	
PN	US5624818-A.	
PD	29-APR-1997.	
PF	01-JUN-1994; 252966.	
PR	01-JUN-1994; US-252966.	
PR	19-SEP-1991; US-736195.	
PR	23-JUN-1992; US-903710.	
PR	01-APR-1994; US-226268.	
PA	(HUTC-) HUTCHINSON CANCER RES CENT FRED.	
PI	Ayer DE, Eisenman RN.	
DR	WPJ. 97-258216/23.	
DR	N-PDSB; T70133.	
PT	msln nucleic acids encoding recombinant polypeptide(s) that	
PT	associate with Mad polypeptide - are possible homologues of S.	
PT	cerevisiae general repressor protein	
PS	Example 2; Fig 2C; 11pp; English.	
CC	This sequence represents Max 11, a basic helix-loop-helix zipper	
CC	(bHLHZip) protein. Max 11 and Max 14 cDNAs appear to be partial,	
CC	overlapping cDNAs. Subsequent isolation of several overlapping cDNAs	
CC	from a Manca (human Burkitt's lymphoma cell line) lambda gt10 library	
CC	permitted deduction of an apparently complete open reading frame for	
CC	Max that encodes 151 residues. The 9-amino acid insertion found in	
CC	several PCR clones is not indicated in the specification. Max is an	
CC	obligate partner for the DNA binding and transcriptional functions of	
CC	MYC family proteins as well as for the Mad protein. Max is a stable,	
CC	ubiquitously expressed protein which in general does not appear to be	
CC	regulated during mitogenesis, the cell cycle, or differentiation.	
CC	Expression of Mad is closely linked to differentiation in at least two	
CC	distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may	
CC	reflect the repression of transcription of Myc regulated genes by Mad.	

CC The DNA, vectors and host cells of the invention are useful for the  
 CC recombinant production of msin proteins useful in elucidation of Mad  
 CC repressor functions.  
 SO Sequence 160 AA:

Query Match 1.9%; Score 149; DB 1; Length 160;  
 Best Local Similarity 29.9%; Pred. No. 1.83e-01;  
 Matches 26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;

Db 7 IEVESDEDEPRFOSAA-D-KRAHNALERRRKHIDKPSLRDVSPLDGEKASRAQIL 64  
 QY 265 MEVQGVKVPINRVQPKVKSANALERRRKHIDKPSLRDVSPLDGEKASRAQIL 323  
 Db 65 DKATEYIOYMRKRNHTHQDIDDLKQ 91  
 QY 324 RKSIDKIRDLQRONHDKAELOQRQRE 350

RESULT 9  
 ID R84882 standard; Protein: 2414 AA.  
 AC R84882;  
 DT 01-FEB-1996 (first entry)  
 DE Transcription factor p300.  
 KW Transcription factor; p300; adenovirus; early region 1A; E1A;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 11..17  
 FT region /label= Nuclear\_location\_signal  
 FT 342..421  
 FT /label= C/H-rich\_region\_1  
 FT /note= "cysteine/histidine-rich region containing  
 FT 2 putative zinc finger motifs"  
 FT 1070..1134  
 FT /label= Bromodomain  
 FT 1162..1461  
 FT /label= C/H-rich\_region\_2  
 FT 1622..1821  
 FT /label= C/H-rich\_region\_3  
 PN W09528499-A1.  
 PD 26-OCT-1995.  
 PF 13-APR-1995; U04682.  
 PR 14-APR-1994; US-227336.  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 PI Eckner R, Ewen M, Livingston D;  
 DR WPI: 95-373813/48.  
 DR N-PSDB: T02792.  
 PT Nucleic acid encoding human p300 that associates with adenovirus E1A  
 PT - and related vectors, host cells and screening assays, also  
 PT diagnosis of cancerous and pre-cancerous tissue by detection of  
 PT mutant p300  
 PS disclosure: Page 61-78; 126pp; English.  
 CC Transcription factor p300 was isolated from human 293 cells and used  
 CC to generate polyclonal antisera in mice. These were used to screen  
 CC 293 CDNA libraries to isolate clones contg. overlapping inserts,  
 CC which were assembled to obtain a full-length CDNA sequence (T02792)  
 CC encoding a protein (R84882) of predicted mol.wt. 264.236 kDa.  
 CC p300 may be produced in host cells (pref. mammalian) and used to  
 CC raise monoclonal antibodies or to screen cpds. for the ability to  
 CC modulate p300-dependent transcription.  
 SO Sequence 2414 AA:

Query Match 1.9%; Score 149; DB 1; Length 2414;  
 Best Local Similarity 24.9%; Pred. No. 1.83e-01;  
 Matches 48; Conservative 50; Mismatches 81; Indels 14; Gaps 11;

Db 818 HCFQQLPQALHONSPPVSRKTPRHHTPPSIGAQQP-ATTIPAVPTPPAMPQPSQ 876  
 QY 120 NCPQOQPTGLKKAQPTATIHMDAQRMPNTAVYPPSLGSSFYQSMSPPTSPYESAQ 179  
 Db 877 ALHP-PP-ROTPPTPTQLDQOV--QPSLPAAPSADQPOQPRSQ--QSTAAVSPTPNAP 930  
 QY 180 NVNVMQPVAAATPAAPASAPLPQOQSYPPFTYNSKAGMTSDAMYLLOPTVAS--PTPSP 238

Db 931 ILPPQATPLSPQAVSIEGVSNPSTSTEVNSQ-AIAEKOPS-OEVKMEAKMEYDQE 988  
 QY 239 VAPPTSTGSRASKVAV-APLA--PSPAMEVGKKVPINRVQPKVKEVRSANALERRY 295  
 Db 989 PADTQPEDISEK 1001  
 QY 296 RTSIN-DKINELK 307

RESULT 10  
 ID W40057 standard; Protein: 2414 AA.  
 AC W40057;  
 DT 20-JUL-1998 (first entry)  
 DE Cellular transcriptional factor p300.  
 KW Cellular transcriptional factor; p300; human; p300; P/CAF;  
 KW transcription; histone acetyltransferase; HIV; infection; cancer;  
 KW therapy; muscle differentiation.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Region 1763..1966  
 FT /note= "P/CAF binding region"  
 PN W09803652-A2.  
 PD 29-JUN-1998.  
 PF 23-JUL-1997; U12877.  
 PR 23-JUL-1996; US-022273.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Howard BH, Nakatani Y;  
 DR WPI: 98-120777/11.  
 DR N-PSDB: V10092.  
 PT New isolated p300/CBP-associated factor, P/CAF - used to develop  
 PT products for modulating transcription, e.g. for treating HIV  
 PT infection or cancers or for promoting muscle differentiation  
 PS disclosure: Page 76-81; 107pp; English.  
 CC This polypeptide sequence comprises p300, a global transcriptional  
 CC coactivator that is involved in the regulation of various  
 CC DNA-binding transcriptional factors. The invention relates to a  
 CC novel human p300/CBP associated cofactor, P/CAF (see W40052), that  
 CC modulates transcription through binding to p300 and CBP (see  
 CC W40058). The region (see W40055) of p300 that binds to P/CAF  
 CC is claimed. The invention provides methods of screening for  
 CC compounds that inhibit or stimulate the transcription modulating  
 CC and histone acetyltransferase activity of P/CAF and p300/CBP.  
 CC Inhibitors can be used e.g. to inhibit HIV Tat-mediated  
 CC transcription in the treatment of HIV infection. Stimulators can  
 CC be used e.g. to activate tumour suppressor p53 in the treatment of  
 CC cancer or to activate the muscle differentiation factor MyoD to  
 CC promote muscle differentiation. The products can also be used to  
 CC inhibit the cell cycle progression inducing effect of an  
 CC oncoprotein which binds p300/CBP in a subject. Also provided is  
 CC a method for determining the amount of P/CAF in a sample by  
 CC contacting the sample with the P/CAF binding region of p300 and  
 CC determining the amount of P/CAF/p300 complex formed.  
 SO Sequence 2414 AA:

Query Match 1.9%; Score 149; DB 1; Length 2414;  
 Best Local Similarity 24.9%; Pred. No. 1.83e-01;  
 Matches 48; Conservative 50; Mismatches 81; Indels 14; Gaps 11;

Db 818 HCFQQLPQALHONSPPVSRKTPRHHTPPSIGAQQP-ATTIPAVPTPPAMPQPSQ 876  
 QY 120 NCPQOQPTGLKKAQPTATIHMDAQRMPNTAVYPPSLGSSFYQSMSPPTSPYESAQ 179  
 Db 877 ALHP-PP-ROTPPTPTQLDQOV--QPSLPAAPSADQPOQPRSQ--QSTAAVSPTPNAP 930  
 QY 180 NVNVMQPVAAATPAAPASAPLPQOQSYPPFTYNSKAGMTSDAMYLLOPTVAS--PTPSP 238  
 Db 931 ILPPQATPLSPQAVSIEGVSNPSTSTEVNSQ-AIAEKOPS-OEVKMEAKMEYDQE 988  
 QY 239 VAPPTSTGSRASKVAV-APLA--PSPAMEVGKKVPINRVQPKVKEVRSANALERRY 295  
 Db 989 PADTQPEDISEK 1001

QY 296 RTSIN-DKINELK 307

RESULT 11  
ID M10042 standard; Protein: 151 AA.  
AC M10042;  
DT 07-FEB-1998 (first entry)  
DE Max 14,  
KW murine; mSinA: mammalian homologue; Saccharomyces cerevisiae: repressor;  
KW Sin3; Mad; Max; mSin:Mad complex; mSin:Mad:Max complex; Myc; promoter;  
KW basic helix-loop-helix zipper protein; compete: DNA-binding;  
KW Myc:Max complex; activate; transcription; gene regulation.  
OS Homo sapiens.  
PN US9624818-A.  
PD 29-APR-1997.  
PF 01-JUN-1994: 252965.  
PR 01-JUN-1994: US-252965.  
PR 19-SEP-1991: US-756195.  
PR 23-JUN-1992: US-903710.  
PR 01-APR-1994: US-222638.  
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PI Ayer DE, Eisenman RN;  
DR WPI: 97-258216/23.  
DR N-PSDB: T70134.  
PT mSin nucleic acids encoding recombinant polypeptide(s) that  
PT associate with Mad polypeptide - are possible homologues of S.  
PT cerevisiae general repressor protein  
PS Example 2; Fig 2D; 11pp; English.  
CC This sequence represents Max 14, a basic helix-loop-helix zipper  
CC (bHLHZip) protein. Max 11 and Max 14 cDNAs appear to be partial,  
CC overlapping cDNAs. Subsequent isolation of several overlapping cDNAs  
CC from a Manca (human Burkitt's lymphoma cell line) lambda gt10 library  
CC permitted deduction of an apparently complete open reading frame for  
CC Max that encodes 151 residues. The 9-amino acid insertion found in  
CC several PCR clones is not indicated in the specification. Max is an  
CC obligate partner for the DNA binding and transcriptional functions of  
CC Myc family proteins as well as for the Mad protein. Max is a stable,  
CC ubiquitously expressed protein which in general does not appear to be  
CC regulated during mitogenesis, the cell cycle, or differentiation.  
CC Expression of Mad is closely linked to differentiation in at least two  
CC distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may  
CC reflect the repression of transcription of Myc regulated genes by Mad.  
CC The DNA, vectors and host cells of the invention are useful for the  
CC recombinant production of mSin proteins useful in elucidation of Mad  
CC repressor functions.  
SQ Sequence 151 AA:

Query Match 1.8%; Score 148; DB 1; Length 151;  
Best Local Similarity 32.4%; Pred. No. 2.13e-01;  
Matches 22; Conservative 23; Mismatches 22; Indels 1; Gaps 1;

Db 15 KRAHNHLEKRRDHKSFSLRDSVPSLOGEKASRAQIIDKATEYIYVRRKHTHTQO 74  
QY 284 KRSAHNALEKRRYRTSINKINELKMLVGEQA-KLKSAAVLKRSIDKTRDLOQNHDLKA 342  
Db 75 DIDLKRO 82  
QY 343 ELQRLQRE 350

RESULT 12  
ID R33386 standard; Protein: 160 AA.  
AC R33386;  
DT 15-JUL-1993 (first entry)  
DE Max protein.  
KW Mad; max; myc; C-myc; helix-loop-helix zipper; leucine zipper;  
KW helix-turn-helix; diagnosis; prognosis; cancer; malignancy;  
KW neoplasm; tumour; studying embryogenesis; study gene regulation.  
FH Key Location/Qualifiers  
FT region 9..17  
FT /note= "9 amino acid insertion found in several  
FT PCR clones."  
FT domain 39..50

FT 1 /label= Helix I  
FT /note= "Helix I of b-HLH homology region."  
FT 57..75  
FT domain /label= Helix II  
FT /note= "Helix II of b-HLH homology region"

PN W09305056-A.  
PD 18-MAR-1993.  
PF 09-SEP-1992: U07629.  
PR 09-SEP-1991: US-756195.  
PR 23-JUN-1992: US-903710.  
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PI Ayer DE, Blackwood EM, Eisenman R;  
DR WPI: 93-100913/12.  
DR N-PSDB: Q38720.  
PT Helix-loop-helix zipper protein named Max - associates with Myc  
PT or Mad polypeptide(s), useful as diagnostic or prognostic tools  
PT for diverse types of cancer  
PS Claim 3; Fig 2; 93pp; English.  
CC This sequence represents the Max protein, and was decoded from the  
CC cDNA isolated as in Q38720. A computer search on a protein database  
CC revealed sequence similarity between a segment of the Max ORF and  
CC the b-HLH proteins, including members of the Myc family. The Max  
CC sequence in this region represents a nearly exact match with the  
CC HLH consensus. This similarity also extends in the amino-terminal  
CC direction into a basic region of Max. The Max sequence just carboxyl  
CC terminal to helix II contains a series of hydrophobic residues,  
CC three of which are leucines, spaced seven residues apart. Helical wheel  
CC analysis of this region suggests that the amphipathic helix II may  
CC extend into and beyond the three leucines. These leucines and the  
CC other non-polar residues might form a hydrophobic face similar to  
CC that in the leucine zipper proteins. It was also shown that a 15  
CC amino acid deletion of the basic region abolishes the capacity of  
CC c-myc to co-transform Rat-1 cells in collaboration with the bcr-abl  
CC oncogene.  
SQ Sequence 160 AA:

Query Match 1.7%; Score 140; DB 1; Length 160;  
Best Local Similarity 32.4%; Pred. No. 6.99e-01;  
Matches 22; Conservative 22; Mismatches 23; Indels 1; Gaps 1;

Db 24 KRAHNHLEKRRDHKSFSLRDSVPSLOGEKASRAQIIDKATEYIYVRRKHTHTQO 83  
QY 284 KRSAHNALEKRRYRTSINKINELKMLVGEQA-KLKSAAVLKRSIDKTRDLOQNHDLKA 342  
Db 84 DIDLKRO 91  
QY 343 ELQRLQRE 350

RESULT 13  
ID Y07044 standard; Protein: 258 AA.  
AC Y07044;  
DT 02-JUL-1999 (first entry)  
DE Breast cancer associated antigen precursor sequence.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer.  
OS Homo sapiens.  
PN W09904265-A2.  
PD 28-JAN-1999.  
PD 15-JUL-1998: U14679.  
PF 15-JUL-1998: US-102322.  
PR 22-JUN-1998: US-896164.  
PR 17-JUL-1997: US-896164.  
PR 10-OCT-1997: US-061599.  
PR 10-OCT-1997: US-061765.  
PR 10-OCT-1997: US-948705.  
PR 11-OCT-1997: GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI: 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -

PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Disclosure: Page 428-429; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 258 AA;

Query Match 1.7%; Score 135; DB 1; Length 258;  
Best Local Similarity 31.9%; Pred. No. 1.45e+00;  
Matches 23; Conservative 22; Mismatches 23; Indels 4; Gaps 4;

Db 13 DPHGRLETHEOGRIKNA-REASQIEKRRRDKMSFIDELASLYPTCNAMSRKIDKLTIV 71  
OY 266 EVOGKVPINRVOQPKYKEVRSANMAIERRYRTSINDKINELKNLV-V-GEQA-KLNKSAV 322  
Db 72 LMAVOHMKTLR 83  
OY 323 LKRSIDKINDLO 334

RESULT 14  
ID R38751 standard; Protein; 644 AA.  
AC R38751.  
DE 23-DEC-1993 (first entry)  
DE DEL protein.  
KW DEL; snapdragon A. majus; regulation; pigmentation; pattern; ivory;  
KW anthocyanin; del mutation; phenotype; corolla; lobe; transformation;  
KW LC: R-S: R gene family; maize; expression vector; marker.  
OS Antirrhinum majus.  
PN WO9314211-A.  
PD 22-JUL-1993.  
PR 08-JAN-1993; G000019.  
PR 09-JAN-1992; US-818570.  
PA (CARP/) CARPENTER R.  
PA (COEN/) COEN E.  
PA (FREE/) FREEDMAN R.  
PA (GOOD/) GOODRICH J.  
PI Carpenter R, Coen E, Freedman R, Goodrich J;  
DR WPI: 93-243226/30.  
DR N-PSDB: 046554.  
PT DNA mol. encoding DEL protein or homologue - capable of  
PT regulating expression of plant genes involved in pigmentation  
PT biosynthesis for regulating anthocyanin genes and plant  
PT pigmentation  
PS Claim 1: Page 26-29; 61pp; English.  
CC This sequence represents the DEL protein derived from the snapdragon,  
CC Antirrhinum majus, which is involved in pigmentation pattern  
CC genes. The recessive del mutation causes a marked difference in  
CC phenotype, causing corolla tubes to be ivory and the lobes to be  
CC fully pigmented. The DEL protein has a strong homology to the  
CC products of lc and R-S, two members of the R gene family which  
CC controls pigmentation in maize. The DNA sequence encoding this  
CC protein may be used, in an expression vector, to control pigmentation  
CC in a range of plants. The DEL coding region may also be used as a  
CC visible marker for gene expression for the detection of transformed  
CC cells.  
SQ Sequence 644 AA;

Query Match 1.7%; Score 134; DB 1; Length 644;  
Best Local Similarity 31.8%; Pred. No. 1.68e+00;

Matches 21; Conservative 19; Mismatches 24; Indels 2; Gaps 2;  
Db 435 PTADIEDRN-VLSEKREKINERFMILASLVPSG-GKVQVSLDHTIDYLGLEKRV 492  
OY 278 PKVKYRSANMAIERRYRTSINDKINELKNLVYGEQAKLNKSAVLRKSIDKINDLOQN 337  
Db 493 DELESN 498  
OY 338 HDLKA 343

RESULT 15  
ID W52830 standard; Protein; 1004 AA.  
AC W52830.  
DE 09-JUL-1998 (first entry)  
DE A tumour suppressor gene called polyhomeotic 1 (hpl1).  
KW Tumour suppressor gene; polyhomeotic 1; hpl1; human chromosome 12p13;  
KW identification; neoplastic tissue; cellular differentiation; diagnosis;  
KW neoplasia; ss.  
OS Homo sapiens.  
PN WO9807858-A1.  
PD 26-FEB-1998.  
PR 06-FEB-1997; U14866.  
PR 23-AUG-1996; US-036939.  
PR 04-DEC-1996; US-031569.  
PA (CHIR ) CHIRON CORP.  
PI Randazzo F;  
DR WPI: 98-169162/15.  
DR N-PSDB: V21060.  
PT Isolated human polyhomeotic 1 oncogene - used to develop products  
PT for diagnosis and therapy of proliferative and developmental  
PT disorders, e.g. neoplasia, dysplasia or hyperplasia  
PS Claim 1: Pages 28-30; 40pp; English.  
CC The present sequence encodes a novel human tumour suppressor gene  
CC termed polyhomeotic 1 (hpl1). The hpl1 gene maps to human chromosome  
CC 12p13, a region which is frequently lost in non-small cell lung cancer  
CC and breast cancer. A method of identifying neoplastic tissue of a  
CC human comprises comparing the expression of a hpl1 gene in a tissue of  
CC a human suspected of being neoplastic with the expression of a hpl1 gene  
CC in a tissue of the human which is normal. Under-expression of the hpl1  
CC gene identifies the subject as having neoplastic tissue. The hpl1  
CC oncogene functions to suppress neoplasia and dysplastic or hyperplastic  
CC cell growth as well as to induce cellular differentiation. The cDNA,  
CC protein and vectors can be used as diagnostic and therapeutic tools for  
CC proliferative and developmental disorders and to identify a p13 region  
CC of a human chromosome 12. They can be used for the detection, diagnosis  
CC or prognosis of neoplasia or for detecting a genetic predisposition to  
CC neoplasia. They can also be used to treat tumours.  
SQ Sequence 1004 AA;

Query Match 1.7%; Score 133; DB 1; Length 1004;  
Best Local Similarity 20.2%; Pred. No. 1.94e+00;  
Matches 49; Conservative 56; Mismatches 123; Indels 12; Gaps 12;

Db 372 TQIOPHSLLQOQOQITLQOKOVYIOQOIAIHQOQFQHQSOQLTATHTLQALQOQOQOQ 431  
OY 48 APTQWYNMLDEPRRTHTQOQSVDDQPOQSVQEVKSEHSVHI-KE-ELHQOQOQSP 105  
Db 432 QOQOQOQOQOATLTLPAPPOVPPVQOQVPPSQOQOQATLVVQPMQSSPLSPD-A-A 489  
OY 106 LLVYRDPDLIATSYNCPQOQPTGLKAAQPTATIHMDAORPMTAVYPPSLGSSFYVQ 165  
Db 490 PK-PPI-PIOSKPPVAPLIPKPOLGAKKMSAQORPHNIVQVUG-TROPGTAAQALGA 546  
OY 166 SWSPTSPVESANOVVNWQVPAATPAPASAPLPQOQTPQPTITVNSKAGMTSDEAMTLL 225  
Db 547 QUAAVPTSRKPGTGVQSGAHLIASSPSSQAFAQLOECPTLA-PG-MTIAVQGTANV 604  
OY 226 -LQPTVASPTSP-PVAPPTSTGSRASKVRVA-PLASPAMEVQKVPINRVOQPKYKE 282  
Db 605 VK 606  
||

Tue Sep 5 10:30:02 2000

US-09-332-522B-8.rag

Page 9

Oy 283 VK 284

Search completed: Sun Sep 3 12:54:06 2000  
Job time : 54 secs.

**THIS PAGE BLANK (USPTO)**





ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS	REFERENCE	RESULTS
Db	301	DKINELKLVGEBOAKLNKSAVLKRSIDTKRIDQRONHDLKAELOQRELMARODSKVK	360			
Qy	301	DKINELKLVGEBOAKLNKSAVLKRSIDTKRIDQRONHDLKAELOQRELMARODSKVK	360			
Db	361	DLLOJGTPGRASKKRRESSQFTTASGLTPPSSDSDSPSLSPMHSIDSLPPSYGSGTA	420			
Qy	361	DLLOJGTPGRASKKRRESSQFTTASGLTPPSSDSDSPSLSPMHSIDSLPPSYGSGTA	420			
Db	421	SCSSGSSSSNEEPLVYPSSMRGMATHSRGLGCMFMEAILAVNPFKTLQGHYDSNDLG	480			
Qy	421	SCSSGSSSSNEEPLVYPSSMRGMATHSRGLGCMFMEAILAVNPFKTLQGHYDSNDLG	480			
Db	481	DMSGORRLISTYVEEGGFVWMOOSSMTWLNFLMLGCLYKLLVYGDPODADQATACQH	540			
Qy	481	DMSGORRLISTYVEEGGFVWMOOSSMTWLNFLMLGCLYKLLVYGDPODADQATACQH	540			
Db	541	RRADPFYSOGSSQAVYAGYLNCLLHFGSLSPASRLCEYQTTWQFRLFPHRLMLGRVL	600			
Qy	541	RRADPFYSOGSSQAVYAGYLNCLLHFGSLSPASRLCEYQTTWQFRLFPHRLMLGRVL	600			
Db	601	SRRSGLGFSSNAASRKQALASARELALLFNRLNLOJLTGNGSRGDMNGIMMALFASMAAEV	660			
Qy	601	SRRSGLGFSSNAASRKQALASARELALLFNRLNLOJLTGNGSRGDMNGIMMALFASMAAEV	660			
Db	661	ANHLTPRETICLHWNTALRMKRSARKYLOQFFARYYMSAROBEGTTRATEQOELRMA	720			
Qy	661	ANHLTPRETICLHWNTALRMKRSARKYLOQFFARYYMSAROBEGTTRATEQOELRMA	720			
Db	721	FLAYGGRYCATHTFTYDLSGEGDGFETLRNCPDPAHVIKOYREHLFFKSIQCLVGA	780			
Qy	721	FLAYGGRYCATHTFTYDLSGEGDGFETLRNCPDPAHVIKOYREHLFFKSIQCLVGA	780			
Db	781	GHRSGGLPTSSVSGEAEOLQOQOHSGLTVSNVLKYTSLKDTLMADEDEDTNVVMADY	840			
Qy	781	GHRSGGLPTSSVSGEAEOLQOQOHSGLTVSNVLKYTSLKDTLMADEDEDTNVVMADY	840			
Db	841	LETAHWMLGEBTLAEOLYGRITKOMPLOLOQCGENHLPALAHVYLRAXMILLKNNGNAL	900			
Qy	841	LETAHWMLGEBTLAEOLYGRITKOMPLOLOQCGENHLPALAHVYLRAXMILLKNNGNAL	900			
Db	901	DKSLKQVLNLODESSVELOECLTVNRITPAKGKIKLIFOLLTCDWLEETPALMEELHMN	960			
Qy	901	DKSLKQVLNLODESSVELOECLTVNRITPAKGKIKLIFOLLTCDWLEETPALMEELHMN	960			
Db	961	MEDDGFYQVPEGLFEFOTDINSLRNIVENIPNAOSRIYLYEAVCRIMAGASPCPTOOL	1020			
Qy	961	MEDDGFYQVPEGLFEFOTDINSLRNIVENIPNAOSRIYLYEAVCRIMAGASPCPTOOL	1020			
Db	1021	DNSLRSRNASHSIFCGSKDRROONFVGGEEPERASAMYACKYLPALLSSGEPAGMLAE	1080			
Qy	1021	DNSLRSRNASHSIFCGSKDRROONFVGGEEPERASAMYACKYLPALLSSGEPAGMLAE	1080			
Db	1081	AAKTEKVGDKRKLKCYOLMKSLGNGISGVKA	1113			
Qy	1081	AAKTEKVGDKRKLKCYOLMKSLGNGISGVKA	1113			

	##status	preliminary
	##molecule-type	mRNA
	##residues	1-1133 ##label SAT
	##cross-references	GB:009103; NID:9516002; PIDN:AAA20085.1; PID:9516003
	CLASSIFICATION	#superfamily 103; #regulatory element binding protein
	SUMMARY	#length 1133 #molecular-weight 120464 #checksum 4586
Query Match	17.7%;	Score 1418; DB 2; Length 1133;
Best Local Similarity	32.8%;	Pred. No. 1,81e-204;
Matches 324;	Conservative 240;	Mismatches 364; Indels 59; Gaps 52.
Db	174	LPGNQGPSSLSASA-PGVSP-IS-LHPVOSSASQQLPASTAPRTTTSIOIORVP 230
QY	147	MEPNNAVTPSLGSGFTYQSMSPPTSPVESANQNVNMQPAAATPAPASAPL-PQ-QSYR 204
Db	231	VLPQHFIFKADSLTLTYTKTDTGATMTKAGISTLAPGTAVOAGPDLQTVSGGTLATVPL 290
QY	205	--QF-FITYSKAGMT--SDEAMTLLQPTVASPTSPSPVAPPTSTSGSRASKRYVA-PL 258
Db	291	VYDIDKLPRIHRLAAGSKALGSAOSRGEKRTAHNAIEKRYNSIDKLYELKDLVVGTEAK 350
QY	259	APSPAMAEVQGVPINR-V-OPKKEVRSAHNAIEREYRFSINDKINELNLVVGQAK 316
Db	351	LKNSAVLRKADYTFEFLQHSNOKLKOEMAL-RN-AAHKSRLDVLVYASGASG-GRDVA 407
QY	317	LKNSAVLRKSIDKTIDLOROHNDLKAEQRLQRELMARFGSKYVDLLOLGTFRPRASKR 376
Db	408	MGVAPREVVDTR-LTPPSDAGSPQSSPLS-LGSKSGSSGGSDE-PDSPYEDSOYKAQ 464
QY	377	RSSQTFPTTDAGLPFRSDESDPRLSPHSDILSPRPYGGSTASCSSGSSSNEEPLVY 436
Db	465	RLHSHGMLDRSLALCALVFLCTGCPNLPLSFGMGICPSPSSASAHHSQSMLEAESRD 524
QY	437	PSSMGKMTHTSLGCMFMFAILVNPETFLQKHVDSNDLQDM-SGQRRIISTUYV-E 494
Db	525	GSMNFWQLLPRPLVWLANLVLVACALLFFVYGEPTPHRTSPAVHNMRRKQADLDLRG 584
QY	495	GEGFVWVOOSKWTMLNTTLMGCIKVLVYGDQDLADQD-A--YCHNRADPFYSQG 551
Db	585	DEPAQAQDLMLTA-LOALGRPLPTSLNLDACSL--WNILRIHLQRLVWGMVLAGRGLR 641
QY	552	QSSQAYAG-YLNCMLHPELSLIPASLE--CYLQTTWQRLRLFRMLWGRVLSRSGGLE 608
Db	642	KDCGLRMDARASARDALVYHKLHOLHMGKYYTGGHILASLWALSALNLAECAGDAVS-M 700
QY	609	SNASARKOALASARELALLENRLNLOLTGSGSGDMNGIMMAFASMAAEVAAHNLTPR 668
Db	701	ALLALITYAAALRYVTSILPRALH-PLTFEFLSSARQC-LAOSGSVPLAMOWLCHPRGHR 758
QY	669	ETTC-IHWMTALRMKRSAPKMWLOQFFARYTMSRAQOBGRTIRAEQTOELMFAATTAIGYR 727
Db	759	F-FVVG-DMAV-HGARQESLISVAGNPVDPDLAOUTYRLECEHILERRALNCIAQPSPTAD- 814
QY	728	YCAHTVFYTDLSDESGEGFTRLRNPQDPAHAHYIKYRREHLRLKSLDCLVGAHKKSSGL 787
Db	815	GDREFSDALGYQLINRCSDAVGTPA-CSFSVSSM-ASTTGDPVAKWASLTVAVVIHW 872
QY	788	PTSSVSGAEQLOQOQHSGTIVSNLYKTSILKDTIMADEDEBDPTNVWMAVDLETAVHW 847
Db	873	LARDEAERLYPLVERRPVLYQE-TER-PLPKALYS-FKAARALLDHR-K-VEGSPAS 927
QY	848	LGEGDTLEQLGRKOMPDTLOQOGENDHLPKA-IHVAVLRAKMILLKNNGNALDKSLQK 906
Db	928	LA-I-CENASGALRSLAAPTGGSID-KAM-OLLCOLLVARTSMQOQSPASAOVA 983
QY	907	LVNIIIDESSVELQCLTVNRTTKDAKGIKLLFOLLTCMWLETRIALWELEHMMWEDGF 966
Db	984	HSASNGQSASALELGFQODLSLRLAQNRPPAMRVFLTEATARLMAGASPARTHOLL 1043
QY	967	YQVP-G-EV--LE-K-FQTDLSLNAYENI.PNQSRIYELVEAVCRLMAGASPCPTQQL 1020
Db	1044	DRSLRRRAGSGK-GTYAAELPR-PLTWREHTAALLASCTLDPPATLSAPGQOMSLAE 1100

QY 1021 DRLSRNAHSSIFCGSDRQONFVGGERASAMVACKYLPALLSPGERAGMLAE 1080

Db 1101 AARTVEKLGDRHLLDCCOMLRLGG 1127

QY 1081 AAKLEKVGDRKRLKECYQLKMSLGN 1107

RESULT 3

ENTRY A54962 #type complete

TITLE sterol regulatory element binding protein 2 precursor - human

ALTRNAME\_NAMES basic-helix-loop-helix-leucine zipper transcription factor

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 24-Sep-1999

ACCESSIONS A49397: A54962

REFERENCE A49397

#authors Hua, X.; Yokoyama, C.; Wu, J.; Briggs, M.R.; Brown, M.S.; Goldstein, J.L.; Wang, X.

#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:11603-11607

#title SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element.

#cross-references MUID:9408581

#accession A49397

#molecule\_type mRNA

#residues 1-1141 ##label HUA

##cross-references GB:U02031; NID:9451329; PIDN:AA50746.1; PID:9451330

##experimental\_source HeLa cells

##note Sequence extracted from NCBI backbone (NCBIP:140987)

REFERENCE A54962

#authors Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.

#journal Genes Dev. (1994) 8:1910-1919

#title Sterol-resistant transcription in CHO cells caused by gene rearrangement that truncates SREBP-2.

#cross-references MUID:95047343

#contents annotation

COMMENT Under conditions of sterol depletion, this membrane-bound transcription factor is released from extranuclear membranes to activate sterol biosynthesis.

GENETICS

#gene GDB:SREBF2; SREBP2

##cross-references GDB:273625

#map\_position Xq24-Xqter

CLASSIFICATION #superfamily sterol regulatory element binding protein

KEYWORDS DNA binding; membrane protein; transcription factor

SUMMARY #length 1141 #molecular\_weight 123672 #checksum 4172

Query Match 15.6%; Score 1328; DB 2; Length 1141;

Best Local Similarity 30.2%; Pred. No. 2,41e-189;

Matches 320; Conservative 270; Mismatches 405; Indels 66; Gaps 50;

Db 118 PTSYPTTPRATPILOPRPOPOPOPTOLOQOTVMI-TP-TFSTTPQTRIIIOQPL-IVONA 174

QY 74 POSVEQOPHYVSEHSSPVHIEKLHQOQOSPLLYKPDPLIATSYNCPQOQPTGLKAA 133

Db 175 ATRSOVLQPOVQSLVTSQOVPTIQOQ-V-QTVQAOQVNLQTNQNGITQITLAPATVQIVA 232

QY 134 OPTATIHMDAQRPMPNTAVIPPSLGSEFVQSSPSPVESANQNVNMP--VAATP 191

Db 233 APQVQVPLVQPOPIIKTDSLVLTLLKTDGSPVMAVONPALTPITQTAALQVPTLV 292

QY 192 APAAPLPQOSYPOFITYN-SKAGMSDEAMYL-LO-PIVAS-PTSPSPVA-PPPIST 246

Db 293 GSGTILITPMVMGQEKVPIKQVPGVKOLEPPEKGGRTTHIIIEKRYSSINDKITE 352

QY 247 GSRASKYRVAPLAPSPAMEV-QCKVPINRVOQKVKESAHNAIEBRYSINDKINE 305

Db 353 LKDLVMTDAKMHSGLYRKALDIYIKYQOVNHLROENMVL-K-LANQKKLLKGI-DL 409

QY 306 LKNLVGEOALNKSAYLRKSIDKIRDLRONHDKALQRLQRLAMRDGSKYVDLQ 365

Db 410 GSLVDNEVDLKE--D-FNQVLLMSPASDSG-SQAGF-SPYSIDSEP--GSPILL--D 459

QY 366 GTRPGRAKKRRRESSQFTTDTAGLTTPRSDSDLSLPMHDSILPSPYGGSTASCSSG 425

Db 460 DAKYKDEPDSPPVAL-GWVDSRIILCVLPFCSPNPLTSLQMGAGHSDQPHSGG 518

QY 426 SSSNEEPILYPPSSMRGNATHSRGLCMFMFAILAVNPFKFLQGRHDSNDLDGMSQ 485

Db 519 RSVLSFESGGGWFDMMPITLLMLVNGVIVLSVFKLVGHEPVIRPHSSSYTFMRH 578

QY 486 RRIISDYVECEGFVAVQOSSMI-WLNFETLMLGGLVLYKVGDPQL--DAQTD--AYOHR 541

Db 579 KQADIDLARCFPAANAANLOTCLAVLGRALPTSLDLACSL--SWNYIRISLQRLVW 636

QY 542 QRADFYSGQSSQAYGYLNCILMPGILSPASRLF--CYLQTTWQFLRFHF--RL--W 595

Db 637 LKRVFOCRIRATPAPTEGFEDEAKTSADALAVHRLHOLITGKLPAGSCSVHMLC 696

QY 596 L-GRVLSRRSGGLFSNASKQALASARELALLENRLQLOLTGNGSGDN--NGIMMALF 653

Db 697 AVNLAECAEERIPSTLVEIHILTPAAMGILTRCGK-Ig-FLASFYLSRAQSLCGPENSAY 754

QY 654 ASNMAEVANHLLTRFETICIHVMTALRK-RSAKRWLOOFFARYMSARQECRTATE 712

Db 755 P-DSLRWLCHPLGQKFMER--SWSVK-SAAKESLYCAQRNPADPIAQVHOAFCKNLLER 810

QY 713 QTOELRWAFAYGYRYCATVHFYDLSGDBODGFFRLRNPCDPAHVIKOYREHLFLK 772

Db 811 AIESIVRQAKKKGKGDQEESECFSSALEYIKLHSPVDSGVNWSPLSRSSVLSALGP 870

QY 773 SIQCLVGA-GHKSGGLPTSSVSGEADLOQOOSHGIIVSNV-LVYTSILKLTQWADER 830

Db 871 DIORWMTSATTVATISWLOGDAVRSHTFVERIPALF-VTES-PLVKAIFPAC-RAM 927

QY 831 DTNVMWADVETAVHMLGEDTLAEOLYGRIKOMPLOQOCCGDHLPKL-HAVLRK 889

Db 928 HASLPGKAD-QQS-S-FCH--CERASGHMSSLVNSGGTSPALNHVYLITCDLLSL 982

QY 900 MILKKNONALDKLQVNLICDESSVELOECLTVNRITDAKIKILFQLTCDWILET 949

Db 983 RIALMOKOASQAVGEFYHNSAGELAGFORDLCSRLRNASFPAYRKVYLHETAYRLM 1042

QY 950 RTALWELEHNMEDDG-FYQVPEVLEKFFQDLMSLRIVENIINASRIYLVEAYCRLM 1008

Db 1043 AGSPPTRHOLEHSLSRRTQSTKH-GEVDAMP-----GGERATATILACRHLPSFL 1096

QY 1009 AGASPCPTQQLDSLSRNASHSIFGSKDRROONTYGGGERASAMVACKYLPALL 1068

Db 1097 SSPQRAVLLAEARTLEKVGDRRSCNDQOMIVKLGGA 1137

QY 1069 SSPGERAGMLAEAKTLEKVGDRKRLKECYQLKMSLGNIG 1109

RESULT 4

ENTRY B54962 #type complete

TITLE sterol regulatory element binding protein 2 precursor - Chinese hamster

ORGANISM #formal\_name Cricetus griseus #common\_name Chinese hamster

DATE 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 04-Sep-1998

ACCESSIONS B54962

REFERENCE A54962

#authors Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.

#journal Genes Dev. (1994) 8:1910-1919

#title Sterol-resistant transcription in CHO cells caused by gene rearrangement that truncates SREBP-2.

#cross-references MUID:95047343

#accession B54962

#status preliminary

#molecule\_type mRNA

#residues 1-1139 ##label YAN

##cross-references GB:U12330

##note 493 Ser was also found

CLASSIFICATION #superfamily sterol regulatory element binding protein



	#cross-references	MIMD:92123207	stabilize both DNA binding and multimerization.		
	#accession	A42029			
	##status	preliminary			
	##molecule-type	mRNA			
	##residues	1-446 ##label	ROM		
	##cross-references	CB:S76673; NID:g243439; PID:g243440			
	##note	sequence extracted from NCBI backbone (NCBIN:76673, NCBIP:76674)			
SUMMARY	#length	446 #molecular-weight	47891 #checksum	8873	
Query Match	2.4%	Score	192;	DB 2; Length	446;
Best local Similarity	25.7%;	Pred. No.	1,01e-08;		
Matches	64; Conservative	65; Mismatches	100; Indels	20; Gaps	16;
Dd	42	PAQVREVLKVGTHLENPTRYHLQQARROOVKOYLSTTIGPKLASALAPPGPSAQT	101		
Oy	122	POOQPTLLKA-AQ-PYATIHMDAQMPNTNVAVPPLSSFVQSMSPTSP--VES-	176		
Dd	102	PAPETAATPTGSAAPSMAALLTGSSSKELDVDELIELSSSYNDMLSYLPQGT	161		
Oy	177	-ANONNVAMPVATAPAPASAPLPPQGSYPQFTI--TYNSAGM-TS-DEAMVLLPPTVA	221		
Dd	162	GLQPSTLPVSGNLDDVSSGVATPATIVSNSCPALPIKRKEISTEAKALIKEROK	221		
Oy	232	S-PIPS-PAAPPTST-GSRASKVRAPLAPS-PAAM-EVQGKVPINRQPVKEY-KR	285		
Dd	222	DNHNLERRRRFINDRIKELGLTIPLKSNDPEMRWNKGITLKASVDYIRKLQEQQRSK-	280		
Oy	286	SAHNAIERRRYTSINDKNELKNIUV-GEQAFL-NKSAVLKRSIDKIRDLOQNHDLKA	342		
Dd	281	DLESQRSLS	289		
Oy	343	ELORQLREL	351		

```

RESULT      8
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MJD:94348499
#accession
#status
#molecule_type mRNA
#residues 1-419 #label RES
#cross-references EMBL:Z29678; NID:9468496; PID:9468497
GENETICS
#gene
#cross-references GDB:214776; OMIM:156845; OMIM:193510
#map position 3p14.1-3p12
SUMMARY
#length 419 #molecular-weight 46938 #checksum 307

Query Match      2.3%  Score 182; DB 2; Length 419;
Best Local Similarity 40.0%; Pred. No. 1,81e-07;
Matches 32; Conservative 22; Mismatches 21; Indels 5; Gaps 4;

Db 197 PALAKERO-KKDHNHLIERRRRININDRIKELGTLPKSNDDPMRNKGTILKASVDYIR 255
      | : : : | | | | | : : : : | : : : : | : : : : | : : : : |
      275 RVQPKVKEVRSANNAIERRYTSINDKINELKNLVV-GQAVL--NKSAYLRKSIDKIR 331

Db 256 KLOREOORAL-ELENROKTL 274
      | | | : : | | | : : |

```



```

Matches 29: Conservative 19; Mismatches 19; Indels 4; Gaps 3
Db 140 KKDHNHLEERRRRNRINDRIKELGTLPKSSDPEKRWKGTITLAAVDYIRKLQKEORS 199
      | : | | | | | | | | | | | | | : : : | | | | | | | :
QY 284 KRSAHMALEERYRIRISINDKINELNLY--VGE-QAKLNKSAVLKRKSIDKINDLQRONHDL 340
      | : | : | | | |
Db 200 K-DESPQRSL 209
      | : | : | | |
QY 341 KAEFLQRLQEL 351

RESULT 12
ENTRY A35658 #type fragment
TITLE transcription factor TFEB - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change
      10-Feb-1997

ACCESSIONS A35658
REFERENCE A35658
#authors Carr, C.S.: Sharp, P.A.
#journal Mol. Cell. Biol. (1990) 10:4384-4388
#title A helix-loop-helix protein related to the immunoglobulin E
      box-binding proteins.
#cross-references MUID:90316407
#accession A35658
#status Preliminary
#molecule_type mRNA
#residues 1-514 #label CAR
##cross-references GB:M33782
      DNA binding; transcription regulation
KEYWORDS #length 514 #checksum 4667
SUMMARY

```

Query Match	2.0%	Score 162:	DB 2:	Length 514:
Best Local Similarity	23.7%	Pred. No. 4.83e-05:		
Matches	71:	Conservative	88:	Mismatches 115; Indels 26; Gaps 22:
Db	111	GAOGEGERERMOQAAVNHMYMO--00000000LGG-PPTPAINTPVAFQSPVPVGEVLKY	168	
OY	75	GSVEEQPHVKSEHSPVHIKEELHQ0000SPLLVYKPKDPLATS--YNCQ09QTGLKA	122	
Db	169	OSTLENTSTYHLOOSQHKVREVISETFYGNKFAAAHSPQALBNPQPPQGGELDTCC	228	
OY	133	-A-O-PTA-TIHMDAQRMPN-TAVYPPSLGSSFFVY-QSM-SPTSPVESANQWVY	183	
Db	229	PPPLATPLISPMALHIGSNPERELDDVIDINIRLMSIGYINPEKQONTPLSSHL	288	
OY	184	MQPVAAIT-PAPASAPLPOOSTPQOFITYNKAGTSDSEAM-YL-L-LQPTVASFTSPPV	239	
Db	289	NVYSSDPOVTASLGVTS-SSCPADLTOKRELDIAESRALAKERO-KKDNHNLERRRRF	346	
OY	240	APPTSTIGSRASKYRVAPLAPSPAAH-EVQG-KVPIRVQPKYEVKRSNAHNAIERRT	297	
Db	347	NINDRIELGLIPKANDLDVVRNKKGITLKASVDYIRRMOKDLOKR-ELEHNSRLRLEMT	405	
OY	298	SINDKINELKNLV-VGE-QAKLNKSAYLRKRSIDIKIRDLORONHDLKAEQLORDEL-MA	353	
RESULT	13			
ENTRY		T04073	#type complete	
TITLE		Intensifier 1 protein - maize		
ORGANISM		#formal_name Zea mays #common_name maize		
DATE		23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999		
ACCESSIONS		T04073		
REFERENCE		Z15188		
AUTHORS		Butt, F.A.; Butt, B.; Scheffler, B.E.; Blewitt, M.; Wienand, U.; Matz, E.C.		
#journal		Plant Cell (1996) 8:1249-1259		
#title		The maize repressor-like gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing.		
#accession		T04073		
#status		preliminary; translated from GB/EMBL/DBJ		

```

##molecule_type DNA
##residues 1-685 ##label BUR
##cross-references EMBL:057893; NID:g1420923; PID:g1420924
##experimental_source strain W22
GENETICS
#gene
#introns in1
#length 685 #molecular_weight 74698 #checksum 2432
SUMMARY

Query Match 2.0%; Score 159; DB 2; Length 685;
Best Local Similarity 22.9%; Pred. No. 1,039-04;
Matches 36; Conservative 51; Mismatches 62; Indels 8; Gaps 7;

Db 423 APSSSSHSIHNGEVOSSPEPCDGDGEGTSRSRGPVSQTELS-ASHVLKERRRREKLN 481
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 242 PPTSTGSASAKSVRAVRLAPLSPAAAEVCGKVPINRYQ-PKXEVYKRSNNHIEKRYTSIN 300
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 482 EGFAMLSLVFV-TKMDRASIIIGDTIEYKQLRRRIQEL--E-SR-RR-LVGSNOKTTM 535
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 301 DKINELKLVVGEQAKLNKSAYLEKRSIDKTRDIDORHNDHKAELIQRBELMARDGSYK 360
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 536 AQQPPPAASTEEERGROTSGGYIARABGTSRAAE 572
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 361 DLLQUGTRPGASKKRDESSDTFTTDGILPPRSDS 397
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT	14	
ENTRY	A42611	#type complete
TITLE	basic-helix-loop-helix-leucine zipper protein delta Max - human	
ALTERNATE_NAMES	max protein 2	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	04-Mar-1993 #sequence_revision 18-Nov-1994	#text_change
ACCESSIONS	07-May-1999	
REFERENCE	A42611; S33119	
#authors	Maxela, T.P.; Koskinen, P.J.; Vastrik, I.; Altitalo, K.	
#journal	Science (1992) 256:373-377	
#title	Alternative forms of Max as enhancers or suppressors of Myc-ras cotransformation.	
#cross-references	MUJD:92229468	
#accession	A42611	
#status	preliminary	
#molecule_type	mRNA	
#residues	1-103 ##label MAK	
#cross-references	GB:X60287; GB:S95058; NID:g599792; PID:g599793	
#experimental	source erythroleukemic cells HEL	
#note	sequence extracted from NCBI backbone (NCBIN:95058, NCBI9:95059)	
REFERENCE	S33118	
#authors	Vaestrik, I.; Koskinen, P.J.; Altitalo, R.; Maekelelae, T.P.	
#journal	Oncogene (1993) 8:503-507	
#title	Alternative mRNA forms and open reading frames of the max gene.	
#cross-references	MUJD:93149613	
#accession	S33119	
#status	preliminary	
#molecule_type	DNA	
#residues	1-103 ##label VAE	
#cross-references	EMBL:X6867; NID:g34469; PID:g34471	
GENETICS	99/1	
#introns	alternative splicing	
KEYWORDS	#length 103 #molecular-weight 12099 #checksum 6857	
SUMMARY		
Query Match	1.9%; Score 149; DB 2; Length 103;	
Best Local Similarity	29.9%; Pred. No. 1.55e-03;	
Matches	26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;	
Db	7	LEVESDEQJRFQSA-D-KRAHHNALEKRRDHINDSFHSDSPSLDGEKASQAQIL 64
	265	MEYGVKVPVIRHVQKVEKVRSAINALERYRIRSDIKINELNLNVAGEDA-KLNSAVI 323

Db 65 DKATEYIOYMRKKNTHOODIDDLKRQ 91  
QY 324 RKSIDKIRDLOQONHDLKAELOQLQRE 350

## RESULT 15

ENTRY S33120 #type complete  
TITLE max protein 3 - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change  
07-May-1999

## ACCESSIONS

S33120

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

S33118

REFERENCE  
#authors Vaestrik, I.; Koskinen, P.J.; Allitalo, R.; Maekela, T.P.  
#journal Oncogene (1993) 8:503-507  
#title Alternative mRNA forms and open reading frames of the max  
gene.

#cross-references MUID:93149613  
#accession S33120

#status preliminary

#molecule-type DNA

#residues 1-134 #label VAE

##cross-references EMBL:X66867; NID:934469; PID:934472

SUMMARY #length 134 #molecular-weight 15395 #checksum 5928

Query Match 1.9%; Score 149; DB 2; Length 134;  
Best Local Similarity 29.9%; Pred. No. 1.55e-03;  
Matches 26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;

Db 7 TEVESDEQPPROSA-D-KRAHNALEKRRDHDKSPHSLRDSVPSLOGEKAQRAQIL 64

QY 265 MEVOGKVPINRYOPKVEKSAHNAIERRYTSTINDKINELKNLVGEQA-KLNKSAVL 323

Db 65 DKATEYIOYMRKKNTHOODIDDLKRQ 91

QY 324 RKSIDKIRDLOQONHDLKAELOQLQRE 350

Search completed: Sun Sep 3 12:52:55 2000  
Job time : 112 secs.





```

Db 241 PPPTSGSRASRVAPLAPSPAMEVOGVKPVINRQPKVKEKRSANATERRTSIN 300
QY 241 PPPTSGSRASRVAPLAPSPAMEVOGVKPVINRQPKVKEKRSANATERRTSIN 300
Db 301 DKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQRONHDLKAELORLQRELMARGSKYK 360
QY 301 DKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQRONHDLKAELORLQRELMARGSKYK 360
Db 361 DLLOGTGPRGSKRRRESSOTFTDAGLTPRPSDESPTSLSPMHSDLSPPSPGGSTA 420
QY 361 DLLOGTGPRGSKRRRESSOTFTDAGLTPRPSDESPTSLSPMHSDLSPPSPGGSTA 420
Db 421 SCSSGSSSSNEEPLVYPSMRGMATHSRGLCMFMAILVNPFTFLQGHYSDNDLG 480
QY 421 SCSSGSSSSNEEPLVYPSMRGMATHSRGLCMFMAILVNPFTFLQGHYSDNDLG 480
Db 481 DMSGORLLSDVDEGEGRVWQSSWMLNTLMGLCYKLLVYGDPLDQDTAYCOH 540
QY 481 DMSGORLLSDVDEGEGRVWQSSWMLNTLMGLCYKLLVYGDPLDQDTAYCOH 540
Db 541 RORADFYRSOGSSOAYVGYLNCMLMFGLSPASRLECTLOTTWQFLRFLRLMGRVL 600
QY 541 RORADFYRSOGSSOAYVGYLNCMLMFGLSPASRLECTLOTTWQFLRFLRLMGRVL 600
Db 601 SRSRSGLSFNMAASRQALASARELALLENRLNQLDTGNSRGDNGIMMALFASNMAEV 660
QY 601 SRSRSGLSFNMAASRQALASARELALLENRLNQLDTGNSRGDNGIMMALFASNMAEV 660
Db 661 AHNLTLPETTCIHVMTALRKRSAPKWLQOPFARYMSRAOEGRRPATQOTELRNA 720
QY 661 AHNLTLPETTCIHVMTALRKRSAPKWLQOPFARYMSRAOEGRRPATQOTELRNA 720
Db 721 FTAVGYRCATHVFTYDSDSGEODGFTRLNPPCDPAHVTKOYREHLFFSIOCIYGA 780
QY 721 FTAVGYRCATHVFTYDSDSGEODGFTRLNPPCDPAHVTKOYREHLFFSIOCIYGA 780
Db 781 GHKSGGLPTSSVSGEAEOLQOQOHSHTIYSNLYKTSILKDTLMDDEDERDNTVVMADY 840
QY 781 GHKSGGLPTSSVSGEAEOLQOQOHSHTIYSNLYKTSILKDTLMDDEDERDNTVVMADY 840
Db 841 LETAVHMLGDEDTLAEOLYGRKQMPLOOQGGENDHLPKALHAVLRAKMITLKNNGAL 900
QY 841 LETAVHMLGDEDTLAEOLYGRKQMPLOOQGGENDHLPKALHAVLRAKMITLKNNGAL 900
Db 901 DKSUKOLVNIICDESSVELQECTLVNRTIDAKIGIKLFPOLLCDWLETRTALWELEHN 960
QY 901 DKSUKOLVNIICDESSVELQECTLVNRTIDAKIGIKLFPOLLCDWLETRTALWELEHN 960
Db 961 MEDDFYGVPGFVLEKFPOTDLSLNINIVENIPNAOSRIYLYEAVCRMLAGASPCFTQOL 1020
QY 961 MEDDFYGVPGFVLEKFPOTDLSLNINIVENIPNAOSRIYLYEAVCRMLAGASPCFTQOL 1020
Db 1021 DRSLSRNAHSSIFCGSDRROQNVGGEREPASAMYVACKTLPALLSSPGERAGMLAE 1080
QY 1021 DRSLSRNAHSSIFCGSDRROQNVGGEREPASAMYVACKTLPALLSSPGERAGMLAE 1080
Db 1081 AAKTLEKVGDKRKLECYQLMKSLGNGISGYKA 1113
QY 1081 AAKTLEKVGDKRKLECYQLMKSLGNGISGYKA 1113

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99161303.
RA INoue J., SATO R.;
RT "A novel, splicing isoform of mouse sterol regulatory element-binding
protein-1 (SREBP-1).";
RL Biosci. Biotechnol. Biochem. 63:243-245(1999).
DR EMBL: AB017337; BAA74795.1; -.
FT NON TER 1
FT 403
SO SEQUENCE 403 AA; 41018 MW; BB57AA8F CRC32;

Query Match 4.2%; Score 336; DB 11; Length 403;
Best Local Similarity 34.8%; Pred. No. 6,11e-33;
Matches 85; Conservative 58; Mismatches 86; Indels 15; Gaps 12;

Db 153 PTPALHTOVQSLASQQLPAPASAPRTNVTYQVOQVPLVLOPHRTKADSLTLTAVKIDAG 212
QY 170 PTPVESANQVNWQVPAATPAPASAPL-PO-QSYP--Qp-FTYNSKA-GMTSDAM 222
Db 213 ATVTAGISTLAPGTAVAGPLOTLSGTTILATVPLVVDTKLPIHRLAAGSKALGSAQ 272
QY 223 YLLQPTVASTPSPPPVAPPTSTGSRASKVRVA-PLAPSPAMEVOGVKPVINR-V-QPK 279
Db 273 SREKRTAHNAIEKRRSSINDKIVEKDLVYGEAKLNKSAVLKRAIDYIRFLQHSNOK 332
QY 280 VKEYRSAHNAIEERRYPTSINDKINELKMLVYGEQAKLNKSAVLKRSIDKIRDLQRONHD 339
Db 333 LKQENTLT-RS-AHKSLSLDVYS-ACSGSGGTDVSECKKPEVER-LTPPSDAGSP 387
QY 340 LKAELORLQRELMARGSKVLDLQLTGPRGSKRRRESSOTFTDAGLTPPSDES DP 399
Db 388 SSSS 391
QY 400 SLSF 403

RESULT 3
ID 09XN00 PRELIMINARY: PRT: 1113 AA.
AC 09XN00:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE Y47D3B.7 PROTEIN.
GN Y47D3B.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
Rhabditiina; Rhabditoidea; Rhabditiidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA MATTHEWS L.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
SMALDON N., SMITH A., SONNHAMMER E., STRADER R., SULTON J.,
THIERREY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
WATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL: AL031635; CAZ21042.1; -.
SO SEQUENCE 1113 AA; 125156 MW; BA56E60 CRC32;

```



	OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia;
	CC	Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesogriecetus.
RN	[1]	SEQUENCE FROM N.A.
RP	MEDLINE:	98167856.
RA	HODGKINSON C.A., NAKAYAMA A., LI H., SWENSON L.-B., OPDECAMP K.,	
RT	ASHER J.A. JR., ARNHETTER H., GLASER T.;	
RT	"Mutation at the anophthalmic white locus in Syrian hamsters:	
RT	haploinsufficiency in the Mitf gene mimics human Waardenburg syndrome	
RT	type 2.";	
RL	Hum. Mol. Genet. 7:703-708(1998).	
DR	EMBL: AF020900; AAC15952.1; -.	
DR	HSSP: P22415; IAN4.	
PR	PFAM: PF00010; HLH; 1.	
FT	NON_TER	
SO	SEQUENCE	
	377 AA; 41728 MW; 83722925 CRC32;	
Query Match	2.3%; Score 181; DB 11; Length 377;	
Best Local Similarity	38.8%; Pred. No. 1,03e-08;	
Matches	31; Conservative 23; Mismatches 21; Indels 5; Gaps 4;	
Db	155 RALAKERO-KKDNNHLERRRRFINDRIKELGLTLPKSNDPDMRWNGKITLKASVDYIR 213	
Oy	275 RVQPPVKVKKSAHAHAIEERRYRTSINDKINLKNLVV-GEQAKL--KMSAVLRKSIDKIR 331	
Db	214 KLQREOQRAR-DLENROKKL 232	
Oy	332 DLORONHDLMKAELORLOREL 351	
RESULT	7	
ID	014841 PRELIMINARY; PRT; 419 AA.	
AC	014841;	
DT	01-NOV-1996 (TREMBLEl. 01, Created)	
DT	01-NOV-1996 (TREMBLEl. 01, Last sequence update)	
DT	01-NOV-1999 (TREMBLEl. 12, Last annotation update)	
DE	MITF PROTEIN (MICROPHTHALMIA-ASSOCIATED TRANSCRIPTION FACTOR).	
GN	MITF.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia;	
CC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE-SKIN;	
RA	MEDLINE: 94348499.	
RA	SCHIBANA M., PEREZ-JURADO L.A., NAKAYAMA A., HODGKINSON C.A., LI X.,	
RA	SCHEIDER M., MIKI T., FEX J., FRANCKE U., ARNHETTER H.;	
RT	"Cloning of MITF, the human homolog of the mouse microphthalmia gene	
RT	and assignment to chromosome 3p14.1-p12.3.";	
RL	Hum. Mol. Genet. 3:553-557(1994).	
RN	[2]	
RP	SEQUENCE OF 1-11 FROM N.A.}	
RP	MEDLINE: 98160190.	
RA	WATANABE A., TAKEIDA K., PLOPLIS B., TACHIBANA M.;	
RT	"Epistatic relationship between Waardenburg syndrome genes MITF and	
RT	PAX3.";	
RL	Nat. Genet. 18:283-286(1998).	
DR	EMBL: Z29678; CAAB2775.1; -.	
DR	EMBL: AF034755; AAC39639.1; -.	
DR	HSSP: P22415; IAN4.	
PR	PFAM: PF00010; HLH; 1.	
KW	DNA-binding.	
SO	SEQUENCE	
	419 AA; 46938 MW; SCE200FD CRC32;	
Query Match	2.3%; Score 182; DB 4; Length 419;	
Best Local Similarity	40.0%; Pred. No. 7,48e-09;	
Matches	32; Conservative 22; Mismatches 21; Indels 5; Gaps 4;	
Db	197 RALAERO-KKDNNHLERRRRFINDRIKELGTLPKSNDPDMRWNGKITLKASVDYIR 255	
Oy	275 RVQPPVKVKKSAHAHAIEERRYRTSINDKINLKNLVV-GEQAKL--KMSAVLRKSIDKIR 331	
Db	256 KLQREOQRAR-ELENROKKL 274	

[illegible]

RC STRAIN-C57BL/6; TISSUE=HEART;  
RX MEDLINE; 9519171.  
RA STEINERINSON E., MOORE K.J., LAMOREUX M.L., FERRE-D'AMARE A.R.,  
RA BURLEY S.K., SANDERS ZIMRING D.C., SKOW L.C., HODGKINSON C.A.,  
RA ARNHEITER H., COPELAND N.G., JENKINS N.A.; mutations helps explain  
RT "Molecular basis of mouse microphthalmia (mi) mutations helps explain  
RT their developmental and phenotypic consequences."  
RL Nat. Genet. 8:256-263(1994).  
CC -1- FUNCTION: ACTS AS A TRANSCRIPTION FACTOR.  
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH TFE3 OR TFE3 (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN, ISOFORMS 1 AND 2,  
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE  
CC SEQUENCE SHOWN HERE IS THAT OF ISOFORM 1.  
CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT HIGH LEVELS IN THE  
CC HEART, SKIN AND SKELETAL MUSCLE. BARELY DETECTABLE IN OTHER  
CC TISSUES. IN THE EMBRYO, EXPRESSED IN DEVELOPING EYE, EAR, SKIN AND  
CC HEART.  
CC -1- DISEASE: DEFECTS IN THE MITF GENE ARE THE CAUSE OF MICROPTHALMIA,  
CC A CONDITION WHICH IS CHARACTERIZED BY ONE OR MORE OF THE  
CC FOLLOWING: LOSS OF PIGMENTATION, REDUCED EYE SIZE, FAILURE OF  
CC SECONDARY BONE RESORPTION, REDUCED NUMBERS OF MAST CELLS AND EARLY  
CC ONSET OF DEAFNESS, AND WHICH GIVES RISE TO A NUMBER OF DIFFERENT  
CC PHENOTYPES.  
CC -1- SIMILARITY: TO OTHER MEMBERS OF THE MYC FAMILY OF HELIX-LOOP-HELIX  
CC TRANSCRIPTION FACTORS.  
DR EMBL: L22958; AAB4773.1; -  
DR EMBL: Z23056; CAA80600.1; -  
DR EMBL: U19874; AAC52155.1; -  
DR EMBL: U19875; AAC52156.1; -  
DR HSSP: P22415; 1AN4.  
DR MGD; MGI:104554; Mltf.  
DR PFAM; PF00010; HLH; 1.  
KW Nuclear protein; Transcription regulation; Alternative splicing;  
KW DNA-binding; Phosphorylation; Disease mutation;  
KW DOMAIN 201 217 ARG/LYS-RICH (BASIC).  
FT DOMAIN 267 288 LEUCINE-ZIPPER (POTENTIAL).  
FT DNA\_BIND 218 260 HELIX-LOOP-HELIX MOTIF.  
FT VARSPIC 1 66 MISALIGNMENT OF EMBL/GENBANK/DBJ DATABASES.  
FT FT OTPAINVSPTPLPSATQVPEYVK -> \*MEMEEXSYO  
FT (IN ISOFORM 2).  
FT VARIANT 67 202 MISSING (IN MICROPTHALMIA WHITE SPOT).  
FT VARIANT 241 246 MISSING (IN MICROPTHALMIA SPOTTED).  
FT VARIANT 242 267 ACIPTESEARALAKEROKDNHMI -> V (IN  
FT MICROPTHALMIA EYELESS-WHITE).  
FT VARIANT 267 267 I -> N (IN MICROPTHALMIA WHITE).  
FT VARIANT 271 271 R -> K (IN MICROPTHALMIA OAK RIDGE).  
FT VARIANT 272 272 MISSING.  
FT VARIANT 277 277 D -> N (IN MICROPTHALMIA VITILIGO).  
FT VARIANT 318 474 MISSING (IN MICROPTHALMIA CLOUDY-EYED).  
FT SEQUENCE 474 AA; 52991 MW; 8E464430 CRC32;  
Query Match 2.3%; Score 191; DB 11; Length 474;  
Best Local Similarity 38.8%; Pred. No. 1,036-08;  
Matches 31; Conservative 23; Mismatches 21; Indels 5; Gaps 4;  
Db 252 RALAKERO-KKDNMTLERRRNNINDRKELGTLIPKSNPDMMKNGTKILKASVDYIR 310  
OY 275 RVQPKVEKSAHNAHLEKRIKINELKLVV-GEQAKL--NRSVAVLRKSIDKIR 331  
Db 311 KLOREOORAK-DLENROKRL 329  
OY 332 DLORONHDKAELORLOREL 351  
RESULT 10  
ID P91527 PRELIMINARY; PRT: 498 AA.  
AC P91527;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE DE SIMILARITY TO A HELIX-LOOP-HELIX MOTIF.  
GN W02C12.3.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilita;  
OC Rhabdilitina; Rhabdilitidae; Rhabdilitidae; Pelodetinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., RILTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROUT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RA MURRAY J., WOHLDMAN P.;  
RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM H.A.  
RC STRAIN-BRISTOL NZ;  
RA WATERSTON R.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U80815; AAB37997.1; -  
DR HSSP: P36956; AM9.  
DR PFAM; PF00010; HLH; 1.  
DR SEQUENCE 498 AA; 54391 MW; EAE8A806 CRC32;  
Query Match 2.3%; Score 188; DB 5; Length 498;  
Best Local Similarity 39.1%; Pred. No. 1,056-09;  
Matches 36; Conservative 25; Mismatches 26; Indels 5; Gaps 4;  
Db 238 KKDINMTERRRRYINDRKELGOMLPKNTSEDKLKKGTILKASCDYIRYLOKDREOA 297  
OY 284 KRSANMAIERKRYRSTINDKINEL-KNLV--VGEQAKLNKSAVLRKSIDKIRLORO-NHD 339  
Db 298 MKTOOQOKSLSTAHKADRYKLEEMAROG 329  
OY 340 LKAELORLQRELMARD-GSKVKDLQGLTRPG 370  
RESULT 11  
ID O75030 PRELIMINARY; PRT: 520 AA.  
AC O75030;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE A-TYPE MICROPTHALMIA ASSOCIATED TRANSCRIPTION FACTOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY;  
RX MEDLINE; 98321192.  
RA AMAR S., FUSE N., YASUMOTO K., SATO S., YAJIMA I., YAMAMOTO H.,  
RA UDONO T., DURLU Y.K., TAKAI M., TAKAHASHI K., SHIBAHARA S.;  
RT "Identification of a novel isoform of microphthalmia-associated  
RT transcription factor that is enriched in retinal pigment epithelium."  
RL Biochem. Biophys. Res. Commun. 247:710-715(1998).  
DR EMBL: AB006909; BAA32288.1; -  
DR HSSP: P22415; 1AN4.  
DR PFAM; PF00010; HLH; 1.  
DR SEQUENCE 520 AA; 58162 MW; F5D32118 CRC32;  
Query Match 2.3%; Score 182; DB 4; Length 520;

```

Best Local Similarity 40.0%      Pred. No. 7,486-09;
Matches 33; Conservative 22; Mismatches 21; Indels 5; Gaps 4;

Db 298 RALAKERO-KKDNNHILERRRRFINIDRIKELGTILPKSNDPDMRKNGKITLAKSVDIR 356
      | : : : | : | | | | : | : | : | : | : : : | : | : | : | : |
Qy 275 RQPFVFKVEKRSRANHAIRRRYRTSINDXINELKNLVV-GEQAKL--NKSAYLRKSIDIR 331
      | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 357 KLOREORAK--ELENRQKL 375
      | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 332 DLQRNHDLKAELOLRQREL 351

RESULT 12      PRELIMINARY:      PRT: 793 AA.
ID 043019
AC 043019:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE HYPOTHEtical 86.8 KD PROTEIN.
SPSP354.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RC WOOD V., RAJANDEAM M.A., BARRELL B.G., HILBERT H., DUESTERHOEFF A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL0202071; CAA17805.1; -
DR HSSP: P36956; IAN9.
DR PFM; PF0010; HLH; 1.
RW Hypothetical protein.
SQ SEQUENCE 793 AA; 86834 MW; 7C9C33AB CRC32;

Query Match 2.1%      Score 166; DB 3; Length 793;
Best Local Similarity 37.5%      Pred. No. 1,226-06;
Matches 30; Conservative 27; Mismatches 18; Indels 5; Gaps 5;

Db 374 GPSDSSSSSTGVSNDSSPISSATFAIQSNGVEPVGLSTQEONLSPLSRRSAHM 433
      | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 236 SPPVAPPTSTGSRAS-KRVVAPLAPSPA-AMEVGKVPIN-RVQ-PKVKEV-RKSAHA 290
      | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 434 IEKRYRNLDNKIAELRDV 453
      | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 291 IERRYRTSINDXINELKNLV 310

RESULT 13      PRELIMINARY:      PRT: 3247 AA.
ID 065553
AC 065553:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
DE UL36.
GN UL36.
OS Bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-COOPER.
RC SCHWYZER M., VICKER C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.,
RA THIRY E., PACES V.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN-COOPER.
RA SCHWYZER M., STYGER D., VOST B., LOWERY D.E., SIMARD C.,
RA LABOISSIERE S., MISRA V., VICKER C., PACES V.;
RL Vet. Microbiol. 0:0-0(1996).
RN (3)
RP SEQUENCE FROM N.A.

```

[illegible]



**THIS PAGE BLANK (USPTO)**



\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
WSEB (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:46:39 2000; Maspar time 34.35 Seconds

Tabular output not generated.

Title: >US-09-332-522B-8  
Description: (1-1113) from US09332522B.pep  
Perfect score: 8012  
Sequence: 1 MDITLMNLIDAPLDESMDLF.....LKECYQLMKSLGSLGYSVKA 1113

Scoring table: PAM 150  
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 57.220; Variance 124.660; scale 0.459

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1494	18.6	1147	1	SREL_HUMAN STEROL REGULATORY ELEM	1.65e-247
2	1418	17.7	1133	1	SREL_CRIGR STEROL REGULATORY ELEM	6.26e-233
3	1278	16.6	1141	1	SRE2_HUMAN STEROL REGULATORY ELEM	1.05e-215
4	1278	16.0	1139	1	SRE2_CRIGR STEROL REGULATORY ELEM	3.68e-206
5	1234	15.4	1023	1	SREL_RAT STEROL REGULATORY ELEM	8.99e-198
6	192	2.4	446	1	TEF3_MOUSE TRANSCRIPTION FACTOR E	2.41e-10
7	190	2.4	743	1	TEF3_HUMAN TRANSCRIPTION FACTOR E	4.45e-10
8	162	2.0	514	1	TEF3_HUMAN TEF3 PROTEIN (FRAGMENT	3.39e-06
9	149	1.9	160	1	MAX_HUMAN MAX PROTEIN.	1.67e-04
10	149	1.9	160	1	MAX_CHICK MAX PROTEIN.	1.67e-04
11	149	1.9	160	1	MAX_RAT MAX PROTEIN.	1.67e-04
12	149	1.9	201	1	YAMC_SCHPO HYPOTHETICAL 23.0 KDA	1.67e-04
13	151	1.9	474	1	VTP3_TTVIV VIRAL PROTEIN TPX	9.27e-05
14	150	1.9	582	1	MNT_HUMAN MAX BINDING PROTEIN MN	1.24e-04
15	150	1.9	817	1	VPR1_YEAST VERPROLIN.	1.24e-04
16	149	1.9	2414	1	P300_HUMAN E1A-ASSOCIATED PROTEIN	1.67e-04
17	142	1.8	160	1	MAX_MOUSE MAX PROTEIN (MYN PROTE	1.26e-03
18	148	1.8	163	1	MAX_XENLA MAX PROTEIN (MAX).	2.24e-04
19	142	1.8	422	1	MYC_AVIC MYC TRANSFORMING PROTE	1.26e-03
20	148	1.8	591	1	MNT_MOUSE MAX BINDING PROTEIN MN	2.24e-04
21	147	1.8	620	1	EXTN_TOBAC EXTENSIN PRECURSOR (CE	7.13e-04
22	144	1.8	865	1	CPN_DROME CALPHOTIN.	7.13e-04
23	141	1.8	1586	1	SN22_HUMAN POSSIBLE GLOBAL TRANSC	1.68e-03

24	133	1.7	310	1	USF1_HUMAN	UPSTREAM STIMULATORY F	1.57e-02
25	133	1.7	310	1	USF1_RABIT	UPSTREAM STIMULATORY F	1.57e-02
26	133	1.7	310	1	USF1_MOUSE	UPSTREAM STIMULATORY F	1.57e-02
27	136	1.7	364	1	HK61_MESAU	HOMEOBOX PROTEIN NKX-6	6.85e-03
28	136	1.7	365	1	HK61_RAT	HOMEOBOX PROTEIN NKX-6	6.85e-03
29	137	1.7	367	1	HK61_HUMAN	HOMEOBOX PROTEIN NKX-6	5.18e-03
30	133	1.7	378	1	HAIR_DROVI	HAIRY PROTEIN.	1.57e-02
31	134	1.7	422	1	MYC_AVIDM	MYC TRANSFORMING PROTE	1.57e-02
32	133	1.7	434	1	NO75_LUPLU	EARLY NODULIN 75 PROTE	1.57e-02
33	136	1.7	454	1	MYC2_MARMO	N-MYC 2 PROTO-ONCOGENE	6.85e-03
34	133	1.7	454	1	MYC2_SPEBE	N-MYC 2 PROTO-ONCOGENE	1.57e-02
35	135	1.7	464	1	SP62_HUMAN	SPICEOSOME ASSOCIATED	9.03e-03
36	135	1.7	583	1	BMAL_HUMAN	BMAL PROTEIN (BRAIN A	9.03e-03
37	133	1.7	700	1	B1B_DROME	NEUROGENIC PROTEIN BIG	1.57e-02
38	140	1.7	822	1	EP58_HUMAN	EPIDERMAL GROWTH FACTO	2.23e-03
39	135	1.7	822	1	Y539_HUMAN	HYPOTHETICAL PROTEIN K	9.03e-03
40	135	1.7	939	1	ST20_YEAST	SERINE/THREONINE-PROTE	9.03e-03
41	137	1.7	2038	1	FSH_DROME	FEMALE STERILE HOMEOTI	5.18e-03
42	131	1.6	331	1	MA2_MESAU	MYC-ASSOCIATED ZINC FI	2.70e-02
43	132	1.6	423	1	MYC_AVIDME	MYC TRANSFORMING PROTE	2.06e-02
44	132	1.6	429	1	MYC5_RAT	S-MYC PROTO-ONCOGENE P	2.06e-02
45	132	1.6	721	1	YK82_MYCTU	HYPOTHETICAL 73.6 KDA	2.06e-02

ALIGNMENTS

RESULT	ID	STANDARD	PRT	ALIGNMENTS
AC	P36956;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL			
DE	REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).			
GN	SREBP1 OR SREBP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
NC	[1]			
RP	SEQUENCE FROM H.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE: 94006541.			
RA	Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,			
RA	Goldstein J.L., Brown M.S.;			
RT	SREBP-1, a basic-helix-loop-helix-leucine zipper protein that			
RT	controls transcription of the low density lipoprotein receptor			
RT	gene.;			
RL	Cell 75:187-197(1993).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 319-394.			
RX	MEDLINE: 98298442.			
RA	Parraga A., Belloso J.L., Ferre-D'Amare A.R., Burley S.K.;			
RT	"Co-crystal structure of sterol regulatory element binding protein 1a			
RT	at 2.3-A resolution."			
RL	Structure 6:661-672(1998).			
CC	-1- FUNCTION: "TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL			
CC	REGULATORY ELEMENT 1 (SRE-1) (5'-ATCAGCCGAC-3') FOUND IN THE			
CC	FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.			
CC	ADL/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO			
CC	BOTH AN E-BOX MOTIF (ATCAGCGA) AND TO SRE-1.			
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER			
CC	B1H PROTEIN.			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR			
CC	ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON			
CC	PROTEOLYTIC CLEAVAGE.			
CC	-1- ALTERNATIVE PRODUCTS: SEVERAL ISOFORMS DIFFERING IN THE N- OR C-			
CC	TERMINI ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN			
CC	IS THAT OF SREBP-1A.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES. MOST			
CC	ABUNDANT IN LIVER AND ADRENAL GLAND. IN FETAL TISSUES LUNG AND			
CC	LIVER SHOWS HIGHEST EXPRESSION.			
CC	-1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPs ARE PROTEOLYTICALLY			
CC	CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS			
CC	AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE			



```

CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09103; AAA20085.1; -
DR HSSP: P36955; 1AM9.
DR PRAM: PF0010; HLH; 1.
DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW Endoplasmic reticulum; Activator; Apoptosis.
FT DOMAIN 1 50 TRANSCRIPTIONAL ACTIVATION (ACIDIC).
FT BINDING 61 176 GLY/PRO/SER-RICH.
FT BINDING 91 410 TO STEROL REGULATORY ELEMENT-1
FT (PROBABLE).
FT DNA_BIND 317 330 BASIC DOMAIN.
FT DOMAIN 331 368 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 367 388 LEUCINE-ZIPPER (POTENTIAL).
FT DOMAIN 421 454 GLY/PRO/SER-RICH.
FT SITE 452 453 CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY
FT SIMILARITY).
FT TRANSMEM 477 497 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
SQ SEQUENCE 1133 AA; 120465 MW; 01A7B09DEDC0A84 CRC64;

Query Match 17.7%; Score 1418; DB 1; Length 1133;
Best Local Similarity 32.8%; Pred. No. 6,266-223;
Matches 324; Conservative 240; Mismatches 364; Indels 59; Gaps 52;

D 174 LPTGTPQPPSSLSASA-PCVSP-IS-LHTOVQSSASQPPAPASTARPTTYSQIQVNP 230
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 147 MPTFAVYPPSLGSEFYQSSPPTSPVESANQVNVVPAATPPASAPL-PQ-QSPY 204
D 231 VLDPPHFKAADSLTLTKTDTGATMKTAGTSTLAPGAVAGPLOTIVSGTILATVPL 290
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 205 ---QP-FTYNSKAGMT-SDEAMYLLOPTYASPTSPVAPPTSTGSRASKRYVA-PL 258
D 291 VVDIDKPIHRLAGSALGSAQSGRGEKRTANAIEKRRSSINDKIYELDLVVGTEAK 350
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 259 APSPAAEVQGVKPINR-V-QPKYKVKRSANALIERRTPTSIDKINELNIVVGEQAK 316
D 351 LNKSAVLRKAIDYTRFLOHSMOKLEQENAL-RN-AAHKSKSLDLYSACCSAG-GTDVA 407
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 317 LNKSAVLRKSIDKIRDLORONHDKAELORELMARDGSKVXDLLOLGTTRPGASKKR 376
D 408 MEGYKPEVVDI-LTPPSPDAGSPSSPLS-IGSRGSSGSDSE-PDSVFEDSOVYKAQ 464
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 377 RESQTTTDDGLTPPSDESDEPSLSPHSDISLPSPYGGSTASCSGSSSNEEPLIV 436
D 465 RLHSHGMLDHSRLALCALVELLCTCNPLASLFGWICPGPSSASAGHHSSGRMSLEASRD 524
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 437 PSSMRGMATHSRLGCMFMFAILAVNPKTFLORGHDSNDLDLM-SGQRRLISYDV-E 494
D 525 GSNKTOWLRLPLVLANGLVLACIALLLFYGEVPTRPHTSPAVHFWHRKROADLCLARG 584
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 495 GEGFAVVOOSWIMLNFITMLGCLVKLKYGPOLDADOT-A--YCOHQRADRYEFSOG 551
D 585 DFNQAAQOQWLIA-LQALGRPLPTSNLDACSL--WNILRLLORLWGRRLAARGLR 641
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 552 QSSQAYVG-YLNCHEMGSLSPASRL--CYLQTTWQRLFLRFLWGRVLSRSGGLE 608
D 642 RDCGLMDARASADALVYHKLHQLHAMGRYTGHLIASMLASLALACADAVS-M 700
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 609 SNASRKQALASARELLLFRLNQDLTGSGSDMMGIMMALFASMAEVAIVNLTPR 668
D 701 ATLAEIVAAALRVKTSIPRALH-FLTRFLISSARQAC-LAOGSGVPLAMQWLCHPVGHR 758
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 669 ETIC-IHVMILRMKRSAPKWLQOEFARVYMSRAROEGRTRATEOTDELFWAFTAYGYR 727

```

```

D 759 F-FVDG-DMAY-HGAPOESLYSVAGNPVPLAOTYRFLCEHLERALNCIAQSPRGAD- 814
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 728 YCAHNVHFTYILSDSGEDGFEFTRLRNPCDPAHAHIKQYREHLRLKSTQCLVGAHKKSSGL 787
D 815 GDREPSDALCYLQLLRNCSDAVGTPA-CSSEVSSEM-ASTTGDPVAKWMASTAVVIHW 872
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 788 PTSSVSGEADLOOQOHHSGITVSVLVKYSTSLTKPDTLWADDEDERDTNVVMADVETAYHW 847
D 873 LRDEEAERLYPLVERPHYLOE-TER-PLPKAIVS-FKAARALLDHR-K-VESGPAS 927
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 848 LLGEDTLAEQLYGRIKQPPLOQCGENDLPAK-LHAIVLAKKILKNNGNALDKSLKQ 906
D 928 LA-I-CEKASGYLRDLSLAAPPTGSSID-KAM-QLLLDLLVARTSMQROOSPASAQVA 983
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 907 LVNLTCDRESSVELOECLVNRITPAKIKLLFOLLTCDWMLLETFTALWELHMMKEDGFE 966
D 984 HSASNGSQAALBLRFGQODLSLRRLAQNFRPMRRVFLHEATARLMAAGSPARTHOLL 1043
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 967 YQVP-G-EV--LE-K-FQTDLNSLRNIVENIPNAQSRITYEAVCAVRLMAGASPCPTOOLL 1020
D 1044 DRSLRRRAGSSGKG-GTVALELRP--PTWRHETAILLASCYLUPAPFLSARPGQOMSLAE 1100
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 1021 DRSLSRRAHSSIFCGSKDRPQONFVGGERASAMTVACKYPLPALISSPERAGMLAE 1080
D 1101 AARTVEKLGDRLLDCCQMLRLRG 1127
O : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Y 1081 AAKLEKVGDKRKLEKCYQLKKSIGNG 1107

RESULT 3
ID SRE2_HUMAN STANDARD; PRT; 1141 AA.
AC Q12772;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-2 (SREBP-2) (STEROL
DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2).
GN SREBP2 OR SREBP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94089681.
RA Hua X., Yokoyama C., Wu J., Briggs M.R., Brown M.S., Goldstein J.L.,
RA Wang X.;
RT "SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that
RT stimulates transcription by binding to a sterol regulatory element.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11603-11607(1993).
RN [2]
RP SEQUENCE OF 91-109.
RX MEDLINE: 94006541.
RA Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,
RA Goldstein J.L., Brown M.S.;
RT "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that
RT controls transcription of the low density lipoprotein receptor
RT gene.";
RL Cell 75:187-197(1993).
RN [3]
RP CLEAVAGE AT ASP-468 BY CASPASES.
RX MEDLINE: 96224103.
RA Pal J.-T., Brown M.S., Goldstein J.L.;
RT "Purification and cDNA cloning of a second apoptosis-related cysteine
RT protease that cleaves and activates sterol regulatory element binding
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:5437-5442(1996).
CC -!- FUNCTION: "TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCAGCCAGC-3') FOUND IN THE
CC FLANKING REGION OF THE LDLR AND HMG-COA SYNTHASE GENES.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR

```

Cc	ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS
Cc	UPON PROTEOLYTIC CLEAVAGE (BY SIMILARITY).
Cc	-I- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN ADULT AND FETAL
Cc	TISSUES.
Cc	-I- PUT: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
Cc	CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
Cc	AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
Cc	PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
Cc	INDEPENDENT OF STEROL LEVELS.
Cc	-I- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLLH) FAMILY OF
Cc	TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
Cc	-----
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc	between the Swiss Institute of Bioinformatics and the EMBL collaboration -
Cc	the European Bioinformatics Institute. There are no restrictions on its
Cc	use by non profit institutions as long as its content is in no way
Cc	modified and this statement is not removed. Usage by and for commercial
Cc	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>
Cc	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).
Cc	-----
Dc	EMBL: U02031; AAA50746.1; -
Dc	HSSP: P36956; IAM9.
Dc	DR HSP: 600481; -
Dc	DR PFAM: PF00010; HLH; 1.
Dc	DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
Kw	Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
Kw	Endoplasmic reticulum.
Ft	DOMAIN 1 50 TRANSRIPTIONAL ACTIVATION (ACIDIC).
Ft	DOMAIN 51 124 GLY/PRO/SER-RICH.
Ft	DOMAIN 125 246 GLY/PRO/SER-RICH.
Ft	DOMAIN 95 423 BINDS TO STEROL REGULATORY ELEMENT-1 (BY
Ft	SIMILARITY).
Ft	DNA_BIND 330 343 BASIC DOMAIN.
Ft	DOMAIN 344 381 LECUCINE-ZIPPER (POTENTIAL).
Ft	DOMAIN 380 401 LECUCINE-ZIPPER (POTENTIAL).
Ft	SITE 468 469 CLEANAGE (BY APOPAIN AND CASPASE-7).
Ft	TRANSMEM 482 502 POTENTIAL.
Ft	TRANSMEM 534 554 POTENTIAL.
Ft	POTENTIAL.
Ft	POLY-ALA.
Ft	DOMAIN 591 595
So	SEQUENCE 1141 AA; 123673 MW; DE42B9CC16CD832B CRC64;
	Query Match 16.6%; Score 1328; DB 1; Length 1141;
	Best Local Similarity 30.2%; Pred. No. 1,05e-215;
	Matches 320; Conservative; 270; Mismatches 405; Indels 66; Gaps 50;
Dd	118 PTSPVTPPRATRIILORRPOPOROPOROLOOQGYMI-T-P-TFSITPQTIRLILOQL-LYQNA 174
Oy	74 POSVEGQRPHVKSEHSHSPVHIKEELHQOOQSFLVTKPPRLATSTNCPOQOPTGLKKA 133
Dd	175 ATSFVLOPVQSVLTYSQGVPTTIOO-V-QTVQAQVVLDTANGTLDTLPATVQVA 232
Oy	134 QPTATHHMDAQRMHPNTAVVPPISSGFVSQDSMSPRTSPVSANQNVMQP--VAAMP 191
Dd	233 APVOQOVPRVLVPOQIIKTDSLVTLLIKTGSPYMAAIVNRPALTALTPTIOTALOFTLV 292
Oy	192 AASAPRLPOQSYROPFITYN-SKAGTSISEAWYL-LQ-PIYAS-PTPBPVA-PEPTST 246
Dd	293 GSGGITLTMPMMGOEKVPIKQVPGVQOLEPPKRGERTNNIEKRYSINDKIIE 352
Oy	247 GSRASKVRKAPRLAPSAAAMEV-OQAKPIRKVQPKYEVKRSAHNALETERTYSINDKIVE 305
Dd	353 LKDLVWGTDAKMHKSGVLKKRAIDYIKYLDQVNHKLQENMYL-K-LANOKNLKGI-DL 409
Oy	306 LKNLVYGEOAKLNKSIAVLKRSIDKIRDLQRHNHDLAELRIQLRELMDAGSKVNRDLGL 365
Dd	410 GSLVNENDLKIE--D-FNQNVLLMSPPASDSG-SQAGR-SPTSIDSEP-GSPLL--D 459
Oy	366 GTRPRASKRRRESQGTFTTDAGLTRPRDESDDSPSLSPHMSDISLSPSYGSTASCSSG 425
Dd	460 DAKVDDEPRSPVAL-GWVDRSRILLCTVPLCLSPNSPLTSLLOWMGAGDSDQHPSGSG 518
Oy	426 SSSSNDEPLVYSSMKGMATHSRLGLCMFMALIANPRTKTLQKRHTYISNDLDGMSQG 485

```

Db 519 RSVLEFSGSGGMEPMWPTLLMLVNCVIVLSEVFKLTVLGVSEVIPPSSSVTFEHRH 578
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 RILSYDVEGEGFAVWOOSMI -WLNTFTMLGCLVLLIYGDQL--DAQD-ALQCHR 544
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 579 KQADILARGDFAAANAANLOTCLAVLGRALPTSRDLASL--SWNIIRYSLQRLVRW 636
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 542 QRADYFSGGSSQAQVAGLYNCLMNFGLSPASRLR--CYLQTTWQFLRFLH--RL--W 595
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 637 LLKKYFCRRATPATGEFDEDAKTSAADAALAYRLHQLHTITKRLPAGSACSDYHALC 696
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 596 L-GRLYSRRSGGLFESNAASRQOALASARELLLRNLRLNLOTLTNGSRGDM-NEIMALF 653
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 697 AVNLAECAEEKIPPTSLVEHLTLTAAGTKTCGKG-LG-FLASYFLSRAOSLCGPEHSAV 754
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 654 ASNMEEVAHNLLTPRETTICHYMTALRRK-RSARKWLQOFFARYMSBARQECRTATE 712
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 755 P-DSLRLWLCHPLGQKFEWER--SWSVK-SAAKESLYCAQRNPADPIAQVHOAFCKNLLR 810
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 713 QTOELRMAFTVAYGRYCATHTVFTDLSDSGGDGFETFLRLRCPDPAHVIKQYREHLFK 772
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 811 AIESLVKROAKKAGDDEESCEFFSSALEYLLKLSFYDSYGVNSPPLSSVYLSKSLGP 870
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 773 STQCLVGA-GHKSGGLPTSSVSGAEQLOQOOSHGTIVSNV-LKTTSLTKLTWADEDER 830
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 871 DLICWMTSATIVIMSLQGDAAVRSHTFVVERHPALE-VKRS-PLVKAIFAC-RAM 922
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 831 DNNVYMAADVLETAHWMLLEDITLAEOLYGRIKOMPTLOQCGGDHLPAL-HAYVRAR 888
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 928 HASLPGKADG--QOS-S-FCH--CERASGHLWSSLNVSQGTSDPALNHVYOLLTCDLLSL 982
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 890 MLKLNNGNALDKSLQKLVNLTCDSSVELEGCLTVNRITDAKIKILFLPOLTCDWLEET 948
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 983 RFLMKQKASASQAVGETIYHASGAEFLAQFQDGLSLRLHSPRAYRKVFLHATVRLM 1048
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 950 RFLMELHMHNMEDDG-FYQVGEVLEKFDQDLSLRNIYENINMAOSRIYLVEAVQRLM 1008
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1043 AGGSTPTFHOLHSLRRTOSTQKH-GEVAMP-----GOREATPILACRLPLSEFL 1098
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1009 AGASCPPTQQLDLDRLSRKNHSSITFCGSKDRKROONFVYGERERASAVYAKTLPALL 1068
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1097 SSPGQRAVLAEAAATTEKVDRRSCNCCQOMIYLGSGTA 1137
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1065 SSPGRAGMLAEAAKTELEKVDKRLKCYOGLMKSLSGIG 1109
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
ID SRE2_CRIGR STANDARD; PRT; 1139 AA.
AC 060429; 060418; 060428; 060427;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-2 (SREBP-2) (STEROL
DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2).
GN SREBP2 OR SREBP2.
OS Cricoetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
RN [1]
RP SEQUENCE FROM N.A. AND TRUNCATED FORM SRD-1.
RX MEDLINE; 95047343.
RA Yang J., Sato R., Goldstein J.L., Brown M.S.;
RT "Sterol-resistant transcription in CHO cells caused by gene
RT rearrangement that truncates SREBP-2.";
RL Genes Dev. 8:1910-1919(1994).
RN [2]
RP SEQUENCE FROM N.A. (TRUNCATED FORMS SRD-1 TO SRD-3).
RX MEDLINE; 95263566.
RA Yang J., Brown M.S., Ho Y.K., Goldstein J.L.;
RT "Three different rearrangements in a single intron truncate sterol
RT regulatory element binding proteins-2 and produce sterol-resistant
RT phenotypic in three cell lines. Role of introns in protein
RT evolution.";
```



DE DETERMINATION-AND DIFFERENTIATION-DEPENDENT FACTOR 1) (ADD1)  
 DE (FRAGMENT).  
 GN SREBP1 OR SREBP1.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-ADIPOCYTE.  
 RX MEDLINE: 93330269  
 RA Tortonoz P., Kim J.B., Graves R.A., Spiegelman B.M.;  
 RT "ADP1: a novel helix-loop-helix transcription factor associated with  
 RT adipocyte determination and differentiation.";  
 RL Mol. Cell. Biol. 13:4753-4759(1993).  
 RN [2]  
 RP DNA-BINDING.  
 RX MEDLINE: 95257939.  
 RA Kim J.B., Spotts G.D., Halvorsen Y.D., Shih H.M., Ellenberger T.,  
 RA Towle H.C., Spiegelman B.M.;  
 RT "Dual DNA binding specificity of ADP1/SREBP1 controlled by a single  
 RT amino acid in the basic helix-loop-helix domain.";  
 RL Mol. Cell. Biol. 15:2582-2588(1995).  
 CC -1- FUNCTION: REGULATED DURING BOTH ADIPOCYTE DETERMINATION AND  
 CC DIFFERENTIATION. ADP1/SREBP1 HETERODIMER HAS DUAL SEQUENCE  
 CC SPECIFICITY, BINDING TO BOTH AN E-BOX MOTIF (ATCAGCTGA) AND TO  
 CC SRE-1 (5'-ATCAGCCAC-3').  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; SREBP-1A AND SREBP-1C/ADP1  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BROWN ADIPOSE  
 CC TISSUE.  
 CC -1- PUT: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY  
 CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS  
 CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE  
 CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,  
 CC INDEPENDENT OF STEROL LEVELS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION, MANY PROBABLE  
 CC FRAMESHIFTS WERE CORRECTED FROM POSITION 878 ONWARD.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L16995; NOT ANNOTATED CDS.  
 DR PROSITE: PS00038; HELIX-LOOP-HELIX: 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;  
 KW Endoplasmic reticulum; Apoptosis; Alternative splicing.  
 FT BINDING 67 385  
 FT DOMAIN 125 128  
 FT DNA\_BIND 293 306  
 FT DOMAIN 307 344  
 FT DOMAIN 343 364  
 FT SITE 427 428  
 FT TRANSMEM 454 474  
 FT TRANSMEM 513 533  
 FT NON\_TER 1023 1023  
 SO SEQUENCE 1023 AA; 108571 MW; 2593DF46B2A11C8F CRC64;  
 Query Match 15.4%; Score 1234; DB 1; Length 1023;  
 Best Local Similarity 32.1%; Pred. No. 8,99e-198;  
 Matches 290; Conservative 218; Mismatches 340; Indels 56; Gaps 50;

QY 147 MPPTAVYVPSLGSSFFVYQSMSPPTSPVESANOVNMQVAAATPAP-ASAPLPQ-QSYP 204  
 DB 207 VYLOPHFIKADSLTLTAVKTPGTATGATAGINTLAPATAVOAGPLOTVSGTITATVPL 266  
 QY 205 ---QP-FITYSKA-GMTSDAMYLLOPYASPPSPVAPPPISTSRASKRYVA-PL 258  
 DB 267 VVDIDKLPIDHRLAAGKALSAQSGRERTAHNALEKRRSSINDKIYELKDLVVGTEAK 326  
 QY 259 APSPAVEVQGVKPIINR-V-QPKVEYKRSAHNAIERRTSINDKINELKNLVYGEQAK 316  
 DB 327 LNKSAVLKALDIYFLOHSMOKLQENLTL-RS-AKXSLNDLVY-ACSGGGTIVS 382  
 QY 317 LNKSAVLKRSIDKTRDLORHNDKAELORLQRELMADGSKYVDLLOLGRPPGASKR 376  
 DB 383 MEGMKPEYVER-LTPPPDAGSPSSPSPLS-LGSRGSSGSDSPDPAPFEDNOVKQR 440  
 QY 377 RESSOTFTTDAGLTPPRDESDPSLSPHSDISLPSPSYGGSTASCSSGSSSNE-EPVY 435  
 DB 441 LPSHSGMLDPSRLALCYLVELCLTNCPLASLFGWGLTLPDASGVHRSRSGSMLEASR 500  
 QY 436 VPSSRMGMATHSRLGCLGMFAILLAVNPFKFLQGHYDSNDLIG-DMSGGRILSYDV- 493  
 DB 501 DGSNTOULPPLPVWLANGLLVLAALLFYGEPVTRPHSGPAVHFWRKROADLDLAR 560  
 QY 494 EGEGRVAVVQSSWIMLNTFTLMLGCLVRLVYGPDLQARD-A--YCOHQRADPFYSQ 550  
 DB 561 GDPAAQAQWLMA-LQALGRPLTPSLNDLACSLD--NMLVHHLRLORLVWGRMLAQAGL 617  
 QY 551 GOSSQAYVG-YLNCILHMFGLSPASRLD--CYLOTQWFLFLFRILWGLVYSRSSGL 607  
 DB 618 QRDYRLKRDARASADAIVYHKLHQLHAMGKYTGHLVASNLSALNLAECAGDAIS- 676  
 QY 608 FSNASRQALASARELLALFRLNLQLOLTGNSRGDMNGIMMALFASMAEVAHNLTP 667  
 DB 677 MATLAEIYMAALARKTSLPRALH-FLTRFLSSKROAC-LAOSGAVLAQMWLCHPVG 734  
 QY 668 RRTIC-HVMYALNRKRRAKPMLOFFARYYSRROCGRTARQOELRMWAFATAGY 726  
 DB 725 RF-FVDG-DMAV-HGAPESLYSVAGNPVDPLAQVTRFCEHLERALNCIAOPSPGAD 791  
 QY 727 RLCATHVTYDLSDBGDEDFFTRLRNCDDRAHYIKIRREHLFKSKOCLVAGAKHSG 786  
 DB 792 -GHRESDALGYLOLLNSCSDVAGAPA-CSFSVSSM-ATTGTDVPKMWASITAVVH 848  
 QY 787 LPTSSVSGEAELOQOQSGTIVSNVLTYSILKXPTLWAEDEDEPTVYMWADVLETAVH 846  
 DB 849 WLRBEAEARLYRVEHITPOVLOE-TER-PLPRALYS-EKAAALLDHR-K-VESSPA 903  
 QY 847 WLGEDTLAEOLYGRIKOMPLOLOQCGENDHLPKA-LHAVLRAKMWILLKNGNALDKSLK 905  
 DB 904 SLA-I-CERKASGYLRDSTASTASSID-KAM-QLLCDLLIVARTSLXROOASASAGA 959  
 QY 906 QLVNLCDESSVELOECLTVNRITDAKIKILLFOLLTCDWLEETLAL-WELEHNMMD 964  
 DB 960 HGTSGPOASALELRFQHDLSLRRLAOSFRPAMRVFLHEATARLMAAGSPARTHOLL 1019  
 QY 965 -GFYGVV-GEYLE-K-FQDINSLRNIIVENIPNQSRIYLYEAVCRILMAGASPTQOLL 1020  
 DB 1020 DRGI 1023  
 QY 1021 DRSL 1024  
 RESULT 6  
 ID TFE3 MOUSE STANDARD. PRT: 446 AA.  
 AC Q64092;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE TRANSCRIPTION FACTOR E3 (FRAGMENT).  
 GN TFE3 OR TCF3.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```
DR  MIM: 314310;
DR  PFM: PF00010; HLH: 1.
DR  PROSITE: PS00038; HELIX-LOOP-HELIX: 1.
KW  Transcription regulation; DNA-binding; Activator; Nuclear protein;
FT  Chromosomal translocation; Proto-oncogene.
FT  DOMAIN 260
FT  DNAS_BIND 344 359
FT  DOMAIN 360 400
FT  DOMAIN 409 430
FT  DOMAIN 575 743
FT  SITE 178 179
FT  SITE 295
FT  SITE 295
FT  SITE 295
FT  SITE 296
FT  SITE 296
FT  CONFLICT 208 211
FT  CONFLICT 219 219
FT  CONFLICT 222 222
FT  CONFLICT 443 443
FT  CONFLICT 443 443
FT  CONFLICT 455 455
FT  CONFLICT 475 475
FT  CONFLICT 593 593
FT  CONFLICT 557 725
FT  CONFLICT 726 743
FT  CONFLICT 726 743
SQ  SEQUENCE 743 AA: 80007 MW: 1082894553543949 CRC64:
Query Match 2.4%; Score 190; DB 1; Length 743;
Best Local Similarity 25.3%; Pred. No. 4.65e-10;
Matches 63; Conservative 65; Mismatches 101; Indels 20; Gaps 16;
Db 169 PAQVREVLKVTGLHNPRTYHQQARQVQKYLSTITGPTLASQALTPPPGASAPQL 228
OY 122 PQOQPTGLKKA-AQ-PTATIHMDAQRMPNTAVYPPSGSFVYQSMSPSPV--ES- 176
Db 229 PAPEAAHTTGPTGAPNSPFWALTTIGSSKEIDVDIIEILLESYNDMLSTYPGGTT 288
OY 177 -ANQVNWQPVATAPAPASAPLPQOSYQPI--TYNSKAM-TS-DEAMTLLLQPTVA 231
Db 269 GLOPSTLPVSGNLVDVYSSQVATPAIVNSCPALNPRTREISEPAKLKEROK 348
OY 232 S-PTPSP-PVAPPTST-GSRASKVAVAPLAP-PAAM-EVQKVPINVOQKVEV--KR 285
Db 349 DNHNLIERRRRRINDIRKELTLPKSDPEPMNNKGTILKASVDYIRKLQKQORSK- 407
OY 286 SAHNAIERRRIRKINDKINELKNLY--VGE-QAKLNKSAVLKSIDKIDQRONHDLKA 342
Db 408 DLESQRSL 416
OY 343 ELQRLREL 351
RESULT 8
ID TFE3_HUMAN STANDARD: PRT: 514 AA.
AC P19484;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TFE3 PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90318407.
RA Carr C.S., Sharp P.A.;
RT "A helix-loop-helix protein related to the immunoglobulin E box-
RL Mol. Cell. Biol. 10:4384-4388(1990).
CC -!- FUNCTION: THIS PUTATIVE TRANSCRIPTION FACTOR BINDS TO THE
CC -!- USE/MULTI SITE AND PROBABLY RECOGNIZES E-BOX SEQUENCES IN THE
```

```
CC HEAVY-CHAIN IMMUNOGLOBULIN ENHANCER.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33782; AAA36730.1; -.
DR PIR: A35658; A35658.
DR HSSP: P36956; IAM9.
DR TRANSFAC: T00812; -.
DR PFM: PF00010; HLH: 1.
DR PROSITE: PS00038; HELIX-LOOP-HELIX: 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT NON_TER 1 1
FT DOMAIN 1 60
FT DOMAIN 106 140
FT DOMAIN 131 140
FT SIMILAR 240 320
FT DOMAIN 253 262
FT DOMAIN 330 345
FT DOMAIN 346 386
FT DOMAIN 395 416
FT DOMAIN 463 511
FT NON_TER 514 514
SQ SEQUENCE 514 AA: 56687 MW: 840961BEDBF520A9 CRC64:
Query Match 2.0%; Score 162; DB 1; Length 514;
Best Local Similarity 23.7%; Pred. No. 3.39e-06;
Matches 71; Conservative 88; Mismatches 115; Indels 26; Gaps 22;
Db 111 QAQGEQERERQQAQVMTYMQQ-QQQQQQQQLGQ-PPPIAINTPVHPSPPPVREVLKV 168
OY 75 QSVQEQPVHKESHSPVIRKELHQQQQSPLLVYKPPPLATIS--YNCPOQPTGLKKA 132
Db 169 QSYLENPSTYHLSQSOHQKREYLSSETGNKFAAHISPAQALRNPQPPQGEIDTCC 228
OY 133 -A-Q-PTA-TIHMDAQRMPN-TAYVPSLGSFV--QSM-SPTSPVESANQNVN 183
Db 229 PPPLATVLPSPMALHIGSNPERELDVINIKELTWSLOYINDEMOMPTLPLSSHL 288
OY 184 MQPVNAT-PAPASAPLPQOSYQPIFYNSKAGMTSDAM-YL-L-QPTVASPTSPSPV 239
Db 289 NYSSDPQVTASLYGVTS-SSCPADLTQKRELTDASALAKENQ-KKDNNLIERRRRF 346
OY 240 APPPTSGSRASKVAVAPLAPSPAM-EVQ-KVPINVOQKVEVKSANALERRRT 297
Db 347 NINDIRKELGMLIPKANDLVYRWNNKGTILKASVDYIRMKDLOKSR-ELENNHSRLTMT 405
OY 298 SINDKINELKNLY--VGE-QAKLNKSAVLKSIDKIRLOQRONHDLKELQRLREL-MA 353
RESULT 9
ID MAX_HUMAN STANDARD: PRT: 160 AA.
AC P25912; P52163.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAX PROTEIN.
OS Homo sapiens (Human), and Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
```



```

RP SEQUENCE FROM N.A.
RX SPECIES=HUMAN.
RA MEDLINE: 91173286.
RA Blackwood E.M., Eisenman R.N.;
RT "Max: a helix-loop-helix zipper protein that forms a
RL sequence-specific DNA-binding complex with Myc.";
RN Science 251:1211-1217(1991).
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=HUMAN.
RX MEDLINE: 93149613.
RA Vaastrik I., Koskinen P.J., Alltalo R., Maekelae T.P.;
RT "Alternative mRNA forms and open reading frames of the max gene.";
RN Oncogene 8:503-507(1993).
[3]
RN SEQUENCE FROM N.A. (DELTA-MAX).
RP SPECIES=HUMAN.
RX MEDLINE: 92229458.
RA Maekelae T.P., Koskinen P.J., Vaastrik I., Alltalo R.;
RT "Alternative forms of Max as enhancers or suppressors of Myc-ras
RL cotransformation.";
RN Science 256:373-377(1992).
[4]
RN SEQUENCE FROM N.A.
RP SPECIES=F.S.CATUS; TISSUE=THYMUS;
RX MEDLINE: 96063016.
RA Gallagher R.C., Neil J.C., Fulton R.;
RT "Cloning and sequence of the feline max, and max 9 transcripts.";
RN DNA Seq. 5:289-271(1995).
[5]
RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 22-107.
RP SPECIES=HUMAN.
RX MEDLINE: 93241312.
RA Ferre-D'Amaré A.R., Prendergast G.C., Ziff E.B., Burley S.K.;
RT "Recognition by Max of its cognate DNA through a dimeric b/HLH/
RL domain.";
RN Nature 363:38-45(1993).
[6]
RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RP MEDLINE: 97277236.
RA Brownlie P., Ceska T., Lamers M., Romier C., Stier G., Teo H.,
RA Suck D.;
RT "The crystal structure of an intact human Max-DNA complex: new
RL insights into mechanisms of transcriptional control.";
RN Structure 5:509-520(1997).
[7]
RN FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC[GA]TG-3'. THE MYC-MAX COMPLEX IS A
CC TRANSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR.
[8]
RN SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
[9]
RN SUBCELLULAR LOCATION: NUCLEAR.
[10]
RN ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. THE SHORT FORMS DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION.
[11]
RN TISSUE SPECIFICITY: HIGH LEVELS FOUND IN THE BRAIN, HEART AND LUNG
CC WHILE LOWER LEVELS ARE SEEN IN THE LIVER, KIDNEY AND SKELETAL
CC MUSCLE.
[12]
RN PTM: PHOSPHORYLATED (PROBABLE).
[13]
RN SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: M64240; AAA36200.1; .
CC EMBL: M64240; AAA36200.1; .

```

```

DR EMBL: X66867: CAA47338.1; -
DR EMBL: X66867: CAA47338.1; -
DR EMBL: X60887: CAA47338.1; -
DR EMBL: X60887: CAA47338.1; -
DR EMBL: D37786: BAA07038.1; -
DR PIR: S33118: S33118.
DR PDB: 1AN2: 17-SEP-97.
DR PDB: 1HLO: 03-DEC-97.
DR TRANSFAC: T00489; -.
DR TRANSFAC: T01567; -.
DR MIM: 154950; -.
DR PFM: PFO0010: HLH; 1.
DR PROSITE: PS00038: HELIX_LOOP_HELIX; 1.
KW Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation; 3D-structure.
FT DNA_BIND 24 36
FT DOMAIN 39 75
FT DOMAIN 81 102
FT VARSPLIC 13 21
FT VARSPLIC 99 160
FT
FT
FT
SQ SEQUENCE 160 AA; 18275 MW; EB10F3137727A56F CRC64;
Query Match 1.9%; Score 149; DB 1; Length 160;
Best Local Similarity 29.9%; Pred. No. 1.67e-04;
Matches 26; Conservative 28; Mismatches 30; Indels 3; Gaps 3;
Db 7 IEVSEDEQPPFQSA-Q-KRAHNALERKRDRHDKSFSHLSRDSVPSLQCEKASRAOIL 64
QY 265 MEVQCKVPIHNVQPVKVKFSRANHALEFRRYRTSINDKINELKLNVGEQA LKNSAVL 323
Db 65 DKATEYIQMKRRKNTHTHOODIDDLKQ 91
QY 324 RKSIDKIRDLQRNHDLKAEIQLRLORE 350
RESULT 10
ID MAX.CHICK STANDARD: PRT: 160 AA.
AC P52162:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE MAX PROTEIN.
GN MAX.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC [1]
RP SEQUENCE FROM H.A.
RC TISSUE-FTIBROBLAST;
RX MEDLINE, 94119609.
RA Sollenberger K.G., Kao T.L., Taparowsky E.J.;
RT "Structural analysis of the chicken max gene.";
RL Oncogene 9:661-664(1994).
CC -1- FUNCTION: "TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC(GA)TG-3'. THE MYC-MAX COMPLEX IS A
CC TRANSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR (BY SIMILARITY)."
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING IN THE SHORT FORMS DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION.
CC -1- PTM: PHOSPHORYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```





AC P37370; Q06133; (Created)  
DI 01-OCT-1994 (Rel. 30, last sequence update)  
DI 01-NOV-1997 (Rel. 35, last sequence update)  
DI 01-NOV-1997 (Rel. 35, last annotation update)  
DE VERPROLIN.  
GN VRP1 OR MDP2 OR ENDS OR YLR337W OR L8300.13.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A364;  
RX MEDLINE; 95058201.  
RA Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;  
RT "A proline-rich protein, verprolin, involved in cytoskeletal  
RT organization and cellular growth in the yeast Saccharomyces  
RT cerevisiae";  
RL Mol. Microbiol. 10:585-596(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
RA Mardis E., Menezes S., Miller N., Nham M., Pauley A., Peluso D.,  
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,  
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
CC -i- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR  
CC GROWTH. MAY EXERT ITS EFFECTS ON THE CYTOSKELETON DIRECTLY, OR  
CC INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR  
CC PROTEINS POSSESSING SH3 DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 226645; CAA81388.1; -;  
DR EMBL; 019028; AAB67263.1; -;  
DR PIR; S39626; S39626.  
DR SGD; L0002481; VRP1.  
KW Cytoskeleton; Repeat.  
FT DOMAIN 5 14  
FT DOMAIN 239 245  
FT DOMAIN 349 357  
FT DOMAIN 396 406  
FT DOMAIN 424 431  
FT DOMAIN 462 468  
FT DOMAIN 704 708  
FT DOMAIN 308 308  
FT CONFLICT 350 350  
FT CONFLICT 689 689  
FT CONFLICT 710 817  
SQ SEQUENCE 817 AA: 82593 MW: 24675223D5B1CA1C8 CRC64: 1;  
Query Match 1.98; Score 150; DB 1; Length 817;  
Best Local Similarity 25.38; Pred. No. 1.24e-04;  
Matches 46; Conservative 48; Mismatches 81; Indels 7; Gaps 7;

Db 293 SSTRKIOTENHKSPSPPLSSAP-P-IPITSHA-PPLPPTAPPSPSLPNVTSAPKKAATSAP 349  
QY 89 SPVHIKEELHQOQOQSPFLVYKRPDLIATSYNCPQOQPTGLAKAOPATITIHMDAQRMP 148  
Db 350 APPPPPLPAMSSASTNSVATVPPTLAPPLPNTTSVPPNKASSMPAPPPPPPP-PPA 408

QY 149 -ENTAVYPPSLGSSFEYQSMSPPTSP-VESANQNNVVMQPVAAATPAPASAPLPQOSYPOP 206  
Db 409 FSTSSALSSSTIPPLAPLPPPPPSVATSVPSAPPPPPPTLTITTKPPASSQOSKISSSSSS 468  
QY 207 FTTVNS-KAGMTSDXAMILLIQTVAASPPSPVAPPPTSTGSRASRVVAPLAPSPAM 265  
Db 469 AV 470  
QY 266 EV 267

Search completed: Sun Sep 3 12:47:47 2000  
Job time : 68 secs.

\*\*\*\*\*  
[W][O][R][D] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:45:17 2000; Maspar time 32.19 Seconds

Tabular output not generated. 910.330 Million cell updates/sec

Title: >US-09-332-522B-6  
Description: (1-1237) from US09332522B.pep  
Perfect Score: 9226  
Sequence: 1 MKNRKSKHFSKAMKFWPAC.....CDYGNFIRVYRPIVADKCH 1237

Scoring table:  
PAM 150  
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a:geneseq35  
1:geneseqp

Statistics: Mean 40.573; Variance 189.529; scale 0.214

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	271	2.9	1170	1 W88447	Yeast NPC1 protein ort	2.40e-11
2	244	2.5	587	1 Y03204	Amino acid sequence of	3.52e-09
3	237	2.5	1278	1 W88445	Human NPC1 (Niemann-Pi	1.27e-08
4	240	2.5	1296	1 W47157	Nevold basal cell carc	7.32e-09
5	240	2.5	1434	1 W52199	Mouse patched (ptc) pr	7.32e-09
6	240	2.5	1434	1 R94380	Mouse patched protein.	7.32e-09
7	240	2.5	1434	1 W72968	Mouse patched protein.	7.32e-09
8	240	2.5	1447	1 W52200	Human patched (ptc) pr	7.32e-09
9	240	2.5	1447	1 W72969	Human patched protein.	7.32e-09
10	240	2.5	1447	1 R75375	Human patched protein.	7.32e-09
11	218	2.4	1311	1 W72971	Precis coenia patched	3.95e-07
12	224	2.4	1319	1 W88446	Mouse NPC1 orthologue.	1.34e-07
13	215	2.3	1311	1 W52197	Precis coenia (buterif	6.76e-07
14	183	2.0	887	1 R58609	Hamster HMG-CoA reduct	1.91e-04
15	149	1.5	517	1 R85852	WD-40 domain-contg. be	6.02e-02
16	140	1.5	779	1 R85854	WD-40 domain-contg. CD	2.61e-01
17	128	1.4	318	1 R85853	WD-40 domain-contg. GB	1.76e+00
18	131	1.4	906	1 R85853	WD-40 domain-contg. be	1.10e+00
19	121	1.3	1054	1 R58610	Yeast HMG-CoA reductas	5.22e-01
20	116	1.3	1299	1 R86304	Drosophila Patched pro	1.12e+01
21	120	1.3	1545	1 W53602	Rat sulphohylurea rece	6.09e+00
22	118	1.3	3588	1 R34712	Bacillus subtilis strA	8.26e+00
23	114	1.2	240	1 W41601	Secreted protein C195-	1.51e+01

ALIGNMENTS

RESULT	ID	W88447	standard; Protein; 1170 AA.
AC	W88447		
DT	26-APR-1999	(first entry)	
DE	Yeast NPC1 protein orthologue.		
KW	Niemann-Pick disease type C; NPC1 gene; yeast; orthologue;		
OS	diagnosis; therapy; animal model; cholesterol; neurodegeneration.		
PN	Saccharomyces cerevisiae.		
PM	W09901555-A1.		
PD	14-JAN-1999.		
PF	03-JUL-1998; U13862.		
PR	03-JUL-1997; US-051682.		
PA	(USSH ) US DEPT HEALTH & HUMAN RESOURCES.		
PI	Carstee ED, Gu J, Loftus SK, Morris JA, Pavan WJ,		
P1	Penchev PG, Rosenfield MA, Tagle DA;		
DR	WPI; 99-106056/09.		
DR	N-PSDB; X06875.		
PT	New isolated gene, NPC-1 - is associated with Niemann-Pick type C		
PT	disease, used to develop products for the study, diagnosis and		
PT	therapy of the disease		
PS	Disclosure, Page 75-79; 101pp; English.		
CC	This polypeptide comprises the Saccharomyces cerevisiae orthologue		
CC	of the human NPC1 polypeptide (see W88445) that is associated with		
CC	Niemann-Pick disease type 2 (NP-C). The polypeptide shows		
CC	extensive identity (34%) and similarity (57%) to the human		
CC	NPC1 protein. Biochemical and genetic analysis of yeast, worm and		
CC	murine NPC1 model systems will provide resources for understanding		
CC	the role of NPC1 in intracellular cholesterol homeostasis and in		
CC	the aetiology of neurodegeneration in NP-C disease. The provision		
CC	of a human NPC1 cDNA sequence (see X06873) enables methods of		
CC	detecting the presence of mutations in the hNPC1 gene, and thereby		
CC	facilitates the determination of whether an individual is an NP-C		
CC	sufferer or carrier.		
SQ	Sequence 1170 AA;		

Query Match 2.9%; Score 271; DB 1; Length 1170;  
Best local similarity 21.2%; Pred. No. 2.40e-11;  
Matches 36; Conservative 71; Mismatches 57; Indels 6; Gaps 5;

Db	557	DISTVAISYIMKFLVATMALRRKG-KTRLLGSLGLIVLASIVCAAGFTLFGKSTL	615
Oy	305	ELVPYTVAFMFLVAVYPSVRKIDFRSRFLALCSYITTAGSLAMSGLCFFGLTISL	364
Db	616	IIAEVIFPLLAIGDINFLITHEVDRNCEOKPEPSIDOKIISAGRMSPSLMSLQOT	675
Oy	365	OSKDIFPVLTIVGLGNSLVITKSYVM-DEF--FDVKIRVAQALSKGGMHISKTLTLEI	421



```

PR 16-MAY-1997: US-017906.
PR 17-MAY-1996: US-017906.
PR 21-MAY-1996: AU-000011.
PR 07-JUN-1996: AU-000363.
PR 14-JUN-1996: US-019765.
PA (USSH ) US SEC DEPT HEALTH.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Chandrabaram A, Christiansen J, Dean MF, Gallani MR,
PI Gerrard B, Gillies S, Goldstein A, Hahn H, Holmberg E,
PI Lefell DJ, Negus K, Pressman C, Shanley S, Smyth I,
PI Udden AB, Vorechovsky I, Wainwright B, Wicking C,
PI Zaphiropoulos PG, Chevenix-Trench, Toftgard R;
DR WPI: 98-008863/01.
DR N-PSDB: V15949.
PT Nevod basal cell carcinoma syndrome tumour suppressor gene - useful
PS Disclosure: Fig 8: 148pp: English.
CC This is a nevroid basal cell carcinoma syndrome (NBCCS) (PTC) protein.
CC The encoding nucleic acid specifically hybridises, under stringent
CC conditions, to a second nucleic acid consisting of a 6568 (full-length
CC sequence), 1732 (exon 1a, b) (V15998) or 659 (exon 2a) (V15999) base pair
CC sequence, in the presence of a human genomic library. The PTC polypeptide
CC when presented as an antigen elicits the production of an antibody which
CC specifically binds to a polypeptide encoded by the above three sequences.
CC The NBCCS gene and its protein product, is a tumour suppressor, and is a
CC homologue of the Drosophila PATCHED (PTC) gene. Detection of the NBCCS
CC nucleic acid, in particular abnormal sequences, by hybridisation assays
CC is useful for detecting a predisposition to NBCCS or to a basal cell
CC carcinoma (also known as Gorlin syndrome). Alternatively, detection is of
CC the polypeptide and is carried out by immunoassay. Vectors comprising
CC this nucleic acid can be used to treat NBCCS. The PTC polypeptide can
CC mitigate symptoms of NBCCS in an organism. The NBCCS nucleic acid
CC includes one or more mutations, chosen from Exon-5 693nsc, Exon-17
CC 2988del8bp, Exon-21 3538delG, Exon-22 64302n, Exon-12 1711nsc, Exon-12
CC 1659insA, Exon-16 2707delC, and Intron-17 3157-2A to G. The mutation may
CC be a nonsense or frameshift mutation. Frameshift mutations are chosen
CC from 244delCt, 27insA, 464insAC, 693nsc, 804del137, 877delG, 929delC,
CC 1370del176, 1393insTCC, 144del16, 1497dup8, 1639insA, 1711nsc,
CC 2183delTC, 2320insAA, 2392delA, 2574delA, 2583delC, 2596complex,
CC 2707delC, 2748insC, 2749dup7, 2988del8bp, 3014insA, 3352delat and
CC 3538delG. The mutation may be a missense, chosen from G391T, G1148A,
CC G1368A, G1525T, C2050T, C2050T, C3015A, G3193C AND G4302n. Alternatively,
CC the mutation alters mRNA splicing and is chosen from A1055-C, 3157-2A to
CC G and 1493-8ins21. All these mutations are claimed but their sequences
CC are not provided in the specification.
CC Sequence 1296 AA:
SO
Query Match 2.6%; Score 240; DB 1; Length 1296;
Best Local Similarity 22.8%; Pred. No. 7.32e-09;
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;

```

```

PD 04-DEC-1997:
PF 02-JUN-1997: U09553.
PR 31-MAY-1996: US-656055.
PA (REGC ) UNIV CALIFORNIA.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;
DR WPI: 98-032648/03.
DR N-PSDB: V21589.
PT Patched protein other than Drosophila melanogaster patched protein -
PS Used for characterising the phenotype of a tumour
PS Claim 5, Pages 66-71, 86pp: English.
CC This is a mouse patched (ptc) protein. The encoding DNA can be used
CC to construct an expression cassette comprising an altered patch or
CC hedgehog gene. The expression cassette comprises a nucleic acid encoding
CC a patched protein other than a Drosophila melanogaster patched protein,
CC or fragment of at least 12 nucleotides in length, as other than an intact
CC chromosome under transcriptional control of a transcriptional initiation
CC region, and a transcriptional termination region, both functional in an
CC expression host. A genetically engineered mammalian cell comprising this
CC expression cassette as an extrachromosomal element or integrated into the
CC genome of the cell can be predisposed to develop basal cell carcinoma as
CC a result of the transfection. By analysing DNA, functional analysis of
CC patched protein function, or by detecting antibody binding to abnormal
CC patched protein, a genetic predisposition to developmental abnormalities
CC and cancer can be diagnosed. This analysis can also be used for
CC characterising the phenotype of a tumour, particularly a carcinoma,
CC especially a basal cell carcinoma. The methods can also be used for
CC characterising transitional cell carcinoma of the bladder, meningiomas
CC medulloblastomas, etc. The modified cells comprising the expression
CC cassette can be used to determine the role of different exons of the
CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal
CC models created from these cells can be used as animal models for
CC carcinomas of the skin. The patched protein of mosquito, butterfly or
CC beetle or alternatively, a mammalian patched protein of human or mouse
CC can be used to identify ligands or substrates that bind to, modulate, or
CC mimic the action of patched gene. These agents could be used as tumour
CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).
CC Sequence 1434 AA;
SO
Query Match 2.6%; Score 240; DB 1; Length 1434;
Best Local Similarity 22.8%; Pred. No. 7.32e-09;
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;

```

PT used to produce antibodies which detect or inhibit patched protein  
ligand signal transduction in cells  
PS Disclosure: Page 46-51; 70pp; English.  
CC Mouse patched protein (PTC) (R4380) was identified as the product  
of a cDNA clone (p14218) derived from mouse limb bud. It has  
CC about 384 identical amino acids to Drosophila PTC. PTC protein has  
CC been proposed as a receptor for hedgehog protein on basis of genetic  
CC experiments in flies. Murine PTC protein can be obtd. in large  
CC amounts by expression of the cDNA clone in transformed host cells.  
CC It can be used to screen for agonists and antagonists, to isolate  
CC its ligand, partic. Sonic hedgehog, to assay for the transcription  
CC of ptc mRNA and to raise antibodies.  
SQ Sequence 1434 AA;

Query Match 2.68; Score 240; DB 1; Length 1434;

Best Local Similarity	22.8%;	Pred. No. 7.32e-09;							
Matches	38;	Conservative	69;	Mismatches	56;	Indels	4;	Gaps	3;

Db 425 VIRVASGYLLMLAYACTLMRLWDCSKSGCAGVLGAVLVAALSVAAAGLGCSIGISFNAA 484  
:: : ::::| :: : | : :: | : ||| : ::  
Qy 306 LVPYVAFMLVEAFVYSVRKIDVERSRLLALCSVITTAGSLAMSLGLCEFFGLTISLQ 365

```

Db      485  TTQVLPFLALGVVDVFLLAHAFSETGONKRIPEDRTGECCLKRTGASVALTISNVTA 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     366  SKDIPYLVILVGLLENSLVITK--SVSNDETFDVKIRVAQALSNEGWHISKTLTFEIT- 422

```

```

Db      545 FEMALIP-IPALRAFLQAAVVVFNFAMVLIFPAISM DYRE 590
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY     423 ILTIGLAFVPIQLQFCIFAIYGLSDENQLMLLFSTILANNIKRE 469

```

RESULT	7
ID	W72968 standard; Protein; 1434 AA.

DT 26-JAN-1999 (first entry)  
DE Mouse patched protein.  
KW Mouse; patched gene; diagnosis; treatment; developmental disorder;  
KW mouse; patched; induced; fibrosis; etc. entry field not 10000000

KW	sperm production; gene therapy.
OS	Mus sp.
PN	US5837538-A.
PO	17-May-1998

PE 06-OCT-1995; 540406.  
PR 06-OCT-1995; US-540406.  
PR 07-OCT-1994; US-319745.  
(CODE) INVTY PRG AND COMMOD THRUOUT

PI Goodrich LV, Johnson RL, Scott MP;  
DR WPI: 99-023461/02.  
DR N-PSDB: V64092.

PT transformants - used to express poly(peptide(s)), useful for diagnosis and treatment of developmental disorders or cancer, and in healing of injured tissue

CC The present sequence represents the mouse patched (ptc) protein. Cells  
CC containing and expressing the ptc gene are used for the recombinant  
CC production of the protein. These in turn are useful: (1) for generating

CC (potential therapeutic agonists and antagonists). The *ptc* gene, or its fragments, are used to isolate related sequences from other mammals; CC identify mutations (particularly those associated with genetic diseases

CC expression levels in testis (to determine relationship with sperm  
CC production) and to isolate 5'-non-coding sequences (used to study  
CC embryonic development and to provide regulated expression of proteins

CC antisense molecules, and to generate transgenic animals for studies c  
CC embryonic development. Ab are used diagnostically to determine the  
CC p1c protein on cell surfaces and as competitive inhibitors of signal

CC express the ptc protein can be used to promote regrowth and healing of  
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
CC protein expression may be useful in cancer treatment (it may control

CC	Wnt-1 oncogene).
SQ	Sequence 1434 AA;

Query match	2.68;	Score 240;	DB 1;	Length 1434;
-------------	-------	------------	-------	--------------

Best Local Similarity	22.8%;	Pred. No. 7.32e-09;
Matches	38; Conservative	69; Mismatches 56; Indels 4; Gaps 3;

Db 425 VIRVASGYLLMLAYACTMLRPDCSKSGAVGLAVLVALSVAAAGLGCLIGISTNNA 484  
 :: : ::::| :: : | :| : ||| ::::  
 QY 306 LVPYVAFMLVAYVYFSVRKIDVRSRFLALCSVITTAGSLANSLGCEFFGLTSLQ 365

```

Db      485 TTQVLPFLALGVDDVFLLAHAFSEIQGNKRIFPEDRIGCECLKTGASVALTISINVA 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     366 SKDIFPYLVILVGLNSLVITK--SVASMDETVDVKIRAAQALSSEGMHISKLTLLTEIT- 422

```

```

Db      545  FEMAALIP-IPALRAFSLOAAVVVVENFAMVLLIFPAISMDLVIRE 590
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY     423  ILTIGLAFVPIQECIFAIYGLLSDMLQMULESTILAMNIKRT 469

```

RESULT	8
ID	W52200
	standard; Protein; 1447 AA.

DT	25-JUN-1998 (first entry)
DE	Human patched (ptc) protein.
KW	patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter;

OS	Homo sapiens.
PN	W09745541-A2.
PD	04-DEC-1997.
DD	02-JUN-1997.
	TT00EE3

PR 31-MAY-1996; US-656055.  
PA (REGC ) UNIV CALIFORNIA.  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
BT  
Encl: 1 P Goodrich IV Johnson RI Oro A Scott MD.

DR WPI; 98-032648/03.  
N-PSDB; V21590.  
PT patched protein other than *Drosophila melanogaster* patched protein -  
used for characterising the phenotype of a tumour

PS Claim 4, Pages 76-80; 86pp; English.  
CC This is a human patched (pc) protein. The encoding DNA can be used  
CC to construct an expression cassette comprising an altered patch or  
CC homolog gene. The expression cassette comprises a nucleic acid encoding

CC a patched protein other than a *Drosophila* melanogaster patched protein,  
CC or fragment of at least 12 nucleotides in length, as other than an intrac  
CC chromosome under transcriptional control of a transcriptional initiation  
CC region and a transcriptional termination region both functional in an

CC expression host. A genetically engineered mammalian cell comprising this  
CC expression cassette as an extrachromosomal element or integrated into the  
CC genome of the cell can be predisposed to develop basal cell carcinoma as  
a result of the transfection. By analyzing DNA functional analysis of

CC patched protein function, or by detecting antibody binding to abnormal CC patched protein, a genetic predisposition to developmental abnormalities CC and cancer can be diagnosed. This analysis can also be used for CC characterizing the phenotype of a tumour particularly a carcinoma

CC especially a basal cell carcinoma. The methods can also be used for CC characterising transitional cell carcinoma of the bladder, meningiomas CC medulloblastomas, etc. The modified cells comprising the expression CC accepto can be used to determine the role of different axes of the

CC patched gene in oncogenesis, signal transduction, etc. Transgenic animal models created from these cells can be used as animal models for CC carcinomas of the skin. The patched protein of mosquito, butterfly or beetle or "alternatively" a mammalian patched protein of human or mouse

CC can be used to identify ligands or substrates that bind to, modulate, or mimic the action of patched gene. These agents could be used as tumour CC suppressors, cell adhesion promoters (e.g. in wound healing and ageing).

Query Match	2.6%;	Score 240;	DB 1;	Length 1447;
Best Local Similarity	22.8%;	Pred: No. 7.32e-09;		
Matches	20;	Concavations	56;	Indols
	60;	Mismatches	56;	Indols
				4;
				3;

Db 439 VIRVASGYLLMLAYACLTMLRNDCKSQGAVGLAGVLLVALSVAAAGLGICSLIGISFNAA 498



```

OY 306 LVPYVAFMLFAVYVSVKRIDVFRSRFLALCSVITTAGSLAMSLGCLFFFGITLSLQ 365
DB 499 TTVQVLPFLAIGVDVDFLLAHAFSEFGONKRIPEDETRGCLRTGASVALTSISNVA 558
OY 366 SKDIFPVLIVLGVLENSLVITK--SVVSMDETFDVKIRVAQALSKEGWHISKTLTLEIT- 422
DB 559 FFMALLIP-IPALRAFSLQAAYVVVFNFAMVLLFPALISMDLYRRE 604
OY 423 ILTIGLATFVPVIOEFCIFALVGLSDFMQLMFLSTILAMNIKRT 469

```

```

RESULT 9
ID W72969 standard; Protein; 1447 AA.
AC W72969;
DI 26-JAN-1999 (first entry)
DE Human patched protein.
KW Human; patched gene; diagnosis; treatment; developmental disorder;
KW cancer; healing; injured tissue; ptc; spina bifida; Wnt-1 oncogene;
KW sperm production; gene therapy.
OS Homo sapiens.
PN US5837538-A.
PD 17-NOV-1998.
PF 06-OCT-1995; 540406.
PR 06-OCT-1995; US-540406.
PR 07-OCT-1994; US-319745.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
PI WPI: 99-023461/02.
DR N-PSDB: V54093.
PT Nucleic acid encoding vertebrate patched protein and related
PT transformants - used to express poly(peptide(s), useful for
PT diagnosis and treatment of developmental disorders or cancer, and in
PT healing of injured tissue.
PS Claim 23: Column 63-70; 38pp; English.
CC The present sequence represents the human patched (ptc) protein. Cells
CC containing and expressing the ptc gene are used for the recombinant
CC production of the protein. These in turn are useful: (i) for generating
CC antibodies (Ab); and (ii) to screen for specific-binding ligands
CC (potential therapeutic agonists and antagonists). The ptc gene, or its
CC fragments, are used to isolate related sequences from other mammals; to
CC identify mutations (particularly those associated with genetic diseases
CC such as spina bifida and other developmental disorders); to monitor
CC expression levels in testis (to determine relationship with sperm
CC production) and to isolate 5'-non-coding sequences (used to study
CC embryonic development and to provide regulated expression of proteins).
CC The complete gene can be used in gene therapy, including expression of
CC antisense molecules, and to generate transgenic animals for studies of
CC embryonic development. Ab are used diagnostically to determine the
CC ptc protein on cell surfaces and as competitive inhibitors of signal
CC transduction through the ptc ligand. Cells that have been engineered to
CC express the ptc protein can be used to promote regrowth and healing of
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc
CC protein expression may be useful in cancer treatment (it may control the
CC Wnt-1 oncogene).
CC
SQ Sequence 1447 AA:

```

```

Query Match 2.6%; Score 240; DB 1: Length 1447;
Best Local Similarity 22.8%; Pred. No. 7.32e-09;
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;
DB 439 VIRASGYLMLAVACTMLRMDCSKSGAGVGLVLAALVSAAGLCLGISFNAA 498
OY 306 LVPYVAFMLFAVYVSVKRIDVFRSRFLALCSVITTAGSLAMSLGCLFFFGITLSLQ 365
DB 499 TTVQVLPFLAIGVDVDFLLAHAFSEFGONKRIPEDETRGCLRTGASVALTSISNVA 558
OY 366 SKDIFPVLIVLGVLENSLVITK--SVVSMDETFDVKIRVAQALSKEGWHISKTLTLEIT- 422
DB 559 FFMALLIP-IPALRAFSLQAAYVVVFNFAMVLLFPALISMDLYRRE 604
OY 423 ILTIGLATFVPVIOEFCIFALVGLSDFMQLMFLSTILAMNIKRT 469

```

```

RESULT 10
ID R75375 standard; Protein; 1447 AA.
AC R75375;
DI 30-JUL-1996 (first entry)
DE Human patched protein.
KW Human patched protein.
KW patched gene; ptc protein; embryo development; cellular regulation;
KW signal transduction; ligand; antibody; hedgehog protein.
OS Homo sapiens.
PN M09611260-A1.
PD 18-APR-1996.
PF 06-OCT-1995; U13233.
PR 07-OCT-1994; US-319745.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
PI WPI: 96-209842/21.
DR N-PSDB: T14220.
PT DNA encoding patched protein other than Drosophila patched protein -
PT used to produce antibodies which detect or inhibit patched protein
PT ligand signal transduction in cells
PS Example: Page 56-61; 70pp; English.
CC Human patched protein (PTC) (R75375) was identified as the product
CC of a cDNA clone (T14220) derived from human lung. It has 96%
CC identity and 98% similarity to mouse PTC. PTC protein has
CC been proposed as a receptor for hedgehog protein on the basis of
CC genetic experiments in flies. Human PTC protein can be obtd. in
CC large amounts by expression of the cDNA clone in transformed host
CC cells. It can be used to screen for agonists and antagonists, to
CC isolate its ligand, partic. Sonic hedgehog, to assay for the
CC transcription of ptc mRNA and to raise antibodies.
SQ Sequence 1447 AA:

```

```

Query Match 2.6%; Score 240; DB 1: Length 1447;
Best Local Similarity 22.8%; Pred. No. 7.32e-09;
Matches 38; Conservative 69; Mismatches 56; Indels 4; Gaps 3;
DB 439 VIRASGYLMLAVACTMLRMDCSKSGAGVGLVLAALVSAAGLCLGISFNAA 498
OY 306 LVPYVAFMLFAVYVSVKRIDVFRSRFLALCSVITTAGSLAMSLGCLFFFGITLSLQ 365
DB 499 TTVQVLPFLAIGVDVDFLLAHAFSEFGONKRIPEDETRGCLRTGASVALTSISNVA 558
OY 366 SKDIFPVLIVLGVLENSLVITK--SVVSMDETFDVKIRVAQALSKEGWHISKTLTLEIT- 422
DB 559 FFMALLIP-IPALRAFSLQAAYVVVFNFAMVLLFPALISMDLYRRE 604
OY 423 ILTIGLATFVPVIOEFCIFALVGLSDFMQLMFLSTILAMNIKRT 469

```

```

RESULT 11
ID W72971 standard; Protein; 1311 AA.
AC W72971;
DI 26-JAN-1999 (first entry)
DE Precis coena patched amino acid sequence.
KW Patched gene; ptc; diagnosis; treatment; developmental disorder;
KW cancer; healing; injured tissue; spina bifida; Wnt-1 oncogene;
KW sperm production; gene therapy.
OS Precis coena.
FH Key Location/Qualifiers
FT MISC_difference 348
FT MISC_difference 908
FT MISC_difference 908 /note="unspecified"
PN US5837538-A.
PD 17-NOV-1998.
PF 06-OCT-1995; 540406.
PR 06-OCT-1995; US-540406.
PR 07-OCT-1994; US-319745.
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
PI Goodrich LV, Johnson RL, Scott MP;
PI WPI: 99-023461/02.
PT Nucleic acid encoding vertebrate patched protein and related
PT transformants - used to express poly(peptide(s), useful for

```

PT diagnosis and treatment of developmental disorders or cancer, and in  
PT healing of injured tissue  
PS Example; Column 25-32; 38pp; English.  
CC The present invention describes vertebrate and invertebrate patched (ptc)  
CC genes. Cells containing and expressing the ptc gene are used for the  
CC recombinant production of the protein. These in turn are useful: (i) for  
CC generating antibodies (Ab), and (ii) to screen for specific-binding  
CC ligands (potential therapeutic agonists and antagonists). The ptc gene,  
CC or its fragments, are used to isolate related sequences from other  
CC mammals; to identify mutations (particularly those associated with  
CC genetic diseases such as spina bifida and other developmental disorders);  
CC to monitor expression levels in testis (to determine relationship with  
CC sperm production) and to isolate 5'-non-coding sequences (used to study  
CC embryonic development and to provide regulated expression of proteins).  
CC The complete gene can be used in gene therapy, including expression of  
CC antisense molecules, and to generate transgenic animals for studies of  
CC embryonic development. Ab are used diagnostically to determine the  
CC ptc protein on cell surfaces and as competitive inhibitors of signal  
CC transduction through the ptc ligand. Cells that have been engineered to  
CC express the ptc protein can be used to promote regrowth and healing of  
CC damaged tissue (e.g. growth of new teeth) and regulation of the ptc  
CC protein expression may be useful in cancer treatment (it may control the  
CC Wnt-1 oncogene). The present sequence represents Precis coenia  
CC (butterfly) patched amino acid sequence, from the present invention.  
SQ Sequence 1311 AA;

Query Match	2.48;	Score 218;	DB 1;	Length 1311;
Best Local Similarity	23.18;	Pred. No. 3.95e-07;		
Matches	36;	Mismatches 57;	Indels 3;	Gaps 3;

[illegible]

RESULT 12

ID W88446 standard; Protein; 1319 AA.

AC W88446.

DT 26-APR-1999 (first entry)

DE Mouse NPC1 orthologue.

KW Niemann-Pick disease type C; NPC1 gene; mouse; orthologue;

KW diagnosis; therapy; cholesterol; neurodegeneration.

OS Mus sp.

PN W09901555-A1.

PN 14-JAN-1999.

PF 02-JUL-1998; U13862.

PR 03-JUL-1997; US-051682.

PA (USSH ) US DEPT HEALTH & HUMAN RESOURCES.

P1 Carstee ED, Gu J, Loftus SK, Morris JA, Pavan WJ,

P1 Pentchev PG, Rosenfield MA, Tagle DA;

DR WPI: 99-106056/09.

DR N-PSDB: X06874.

PT New isolated gene, NPC-1 - is associated with Niemann-Pick type C

PT disease, used to develop products for the study, diagnosis and

PT therapy of the disease

PS Claim 1; Page 66-70; 101pp; English.

CC This polypeptide comprises the murine orthologue of human NPC-1 (see

CC W99445) that is associated with Niemann-Pick disease type C (NP-C).

CC (The amino acid sequence, predicted from an isolated cDNA clone

CC (see X06874), comprises a putative N-terminal signal peptide

CC followed by a domain that is unique to the NPC1 orthologues (in

CC mouse, human, *Caenorhabditis elegans* and *Saccharomyces cerevisiae*,

CC see W88445-48), and 13 putative transmembrane domains that include

CC a potential sterol-sensing domain. Biochemical and genetic

CC analysis of NPC1 using mouse model systems will provide resources

CC for understanding the role of NPC1 in intracellular cholesterol  
CC homeostasis and in the etiology of neurodegeneration in NP-C  
CC disease. The provision of a human NPC1 cDNA sequence (see X06873)  
CC provides methods of detecting the presence of mutations in the  
CC NPC1 gene, and thereby facilitates the determination of whether an  
CC individual is an NP-C sufferer or carrier.  
SQ Sequence 1319 AA;

Query Match	2.48;	Score 224;	DB 1;	Length 1319;
Best Local Similarity	26.3%;	Pred. No. 1.34e-07;		
Matches	45;	Conservative	61;	Mismatches 54; Indels 11; Gaps 9;

```

Db 667 IISVWMEFLYSIALGHQSCRLTY-DSKISUGIABIILVSSVACSGIFSVMGPLT 725
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 306 LVPYVAFMLV-FMYVVF-SYRKIDVRSRRLALACSVITTAQSLAMSLGACFFEGGLTIS 363
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 726 LIVIEVPELVLANGVONIFLVQTY-QDBERQOEEF-LDOOIGRIUG-EVAPTMFLSSF 782
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 364 LOSKDIPFYILVIGLENSLVITKSVSMDETVKIRVQAOLSKF-GWHISKTL-LT-- 419
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 783 SETSAFFGALSSMPAVHTSFSLFAGMVLIDFLQITCFPSLGLDIDRKE 833
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 420 -EITLTIGLATFVPIQEFCTFIVLSDLSDFMLQMLTSTILAMNKRIE 469
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT	13
ID	W52197 standard; Protein; 1311 AA

DT 25-JUN-1998 (first entry)  
DE Precis coenia (butterfly) patched (ptc) protein.  
RW Patched protein; ptc; cancer; tumour suppressor; cell adhesion promoter.  
KW wound healing; ageing; Precis coenia; butterfly.  
OS Precis coenia.  
PN MO9745541-A2.  
PD 04-DEC-1997.  
PF 02-JUN-1997; U09553.  
PR 31-MAY-1996; US-656055.  
PA (REGC ) UNIV CALIFORNIA.  
PA (STRD ) UNIV IELAND STANFORD JUNIOR.  
PI Epstein E, Goodrich LV, Johnson RL, Oro A, Scott MP;  
DR WPI: 98-032648/03.

Patched protein other than *Drosophila* melanogaster patched protein -  
 PR used for characterizing the phenotype of a tumour  
 PS Claim 2; Pages 52-55; 86pp; English.  
 CC This is a *Pectis coenia* (butterfly) patched (ptc) protein and the  
 CC encoding DNA can be used to construct an expression cassette comprising  
 CC an altered patch or hedgehog gene. The expression cassette comprises a  
 CC nucleic acid encoding a patched protein other than a *Drosophila*  
 CC melanogaster patched protein, or fragment of at least 12 nucleotides in  
 CC length, as other than an intact chromosome under transcriptional control  
 CC of a transcriptional initiation region, and a transcriptional termination  
 CC region, both functional in an expression host. A genetically engineered  
 CC mammalian cell comprising this expression cassette as an extrachromosomal  
 CC element or integrated into the genome of the cell can be predisposed to  
 CC develop basal cell carcinoma as a result of the transfection. By  
 CC analysing DNA, functional analysis of patched protein function, or by  
 CC detecting antibody binding to abnormal patched protein, a genetic  
 CC predisposition to developmental abnormalities and cancer can be  
 CC diagnosed. This analysis can also be used for characterizing the  
 CC phenotype of a tumour, particularly a carcinoma, especially a basal cell  
 CC carcinoma. The methods can also be used for characterizing transitional  
 CC cell carcinoma of the bladder, meningiomas medulloblastomas, etc. The  
 CC modified cells comprising the expression cassette can be used to  
 CC determine the role of different exons of the patched gene in oncogenesis,  
 CC signal transduction, etc. Transgenic animal models created from these  
 CC cells can be used as animal models for carcinomas of the skin. The  
 CC patched protein of mosquito, butterfly or beetle or alternatively, a  
 CC mammalian patched protein of human or mouse can be used to identify  
 CC ligands or substrates that bind to, modulate, or mimic the action of  
 CC patched gene. These agents could be used as tumour suppressors, cell  
 CC adhesion promoters (e.g. in wound healing and ageing).  
 Sequence 1311 AA;



**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 NWSEI (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:43:14 2000; MasPar time 58.87 Seconds  
 Tabular output not generated. 991.194 Million cell updates/sec

Title: >US-09-332-522B-6  
 Description: (1-1237) from US09332522B.pep  
 Perfect Score: 9226  
 Sequence: 1 MNKTKSKHFKAMKFWPAC.....CDYGNIEVVRFPVADKCH 1237

Scoring table: PAM 150  
 Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: pif64  
 1:pif1 2:pif2 3:pif3 4:pif4

Statistics: Mean 55.949; Variance 117.221; scale 0.477

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	271	2.9	1170	2	S52525	probable membrane pro
2	223	2.4	1182	2	T13952	membrane protein pich
3	212	2.3	932	1	A31898	hydroxymethylglutaryl
4	210	2.3	1055	2	T05663	hypothetical protein
5	183	2.0	887	1	RDHYE	hydroxymethylglutaryl
6	183	2.0	887	1	A23586	hydroxymethylglutaryl
7	183	2.0	888	1	RDHYE	hydroxymethylglutaryl
8	178	1.9	856	1	S30338	hydroxymethylglutaryl
9	178	1.9	856	1	A35728	hydroxymethylglutaryl
10	177	1.9	916	2	S32572	hydroxymethylglutaryl
11	177	1.9	1053	2	S72194	hydroxymethylglutaryl
12	154	1.7	701	2	B71215	hypothetical protein
13	155	1.7	701	2	T16607	hypothetical protein
14	145	1.6	257	2	S62507	hypothetical protein
15	150	1.6	308	2	E75029	protein-export membra
16	147	1.6	431	2	S49821	PrU2 protein - Arabid
17	149	1.6	518	2	B48088	beta-transducin repea
18	145	1.6	1015	2	T15830	hypothetical protein
19	146	1.6	1146	2	A55532	myosin-heavy-chain ki
20	136	1.5	425	2	S78258	probable translocator
21	143	1.5	495	2	T04784	hypothetical protein
22	140	1.5	750	1	D69403	conserved hypothetica
23	137	1.5	779	2	S56245	cell division control

RESULT ENTRY	1	134	1.5	888	2	E71280	probable antibiotic t	6.70e-03
25	129	1.4	282	2	D64456	protein-export membra	2.81e-02	
26	130	1.4	290	2	T02300	GTP-binding regulator	2.11e-02	
27	129	1.4	308	2	S74813	hypothetical protein	2.81e-02	
28	128	1.4	318	2	S11904	GTP-binding regulator	3.73e-02	
29	128	1.4	325	2	T06784	GTP-binding protein b	3.73e-02	
30	126	1.4	334	2	T09613	probable GTP-binding	6.52e-02	
31	126	1.4	428	2	B61817	protein RWD - rice	6.52e-02	
32	127	1.4	485	2	S49820	conserved hypothetica	4.93e-02	
33	131	1.4	486	2	S49820	PL1 protein - Arabid	1.59e-02	
34	133	1.4	714	2	T16126	hypothetical protein	8.95e-03	
35	130	1.4	817	2	S11445	probable membrane pro	2.11e-02	
36	131	1.4	906	2	S35312	coatomer complex beta	1.59e-02	
37	131	1.4	906	2	S35342	Golgi-associated part	1.59e-02	
38	125	1.4	952	2	S35348	H+-transporting Atpas	8.62e-02	
39	125	1.4	953	2	T12577	H+-transporting Atpas	8.62e-02	
40	125	1.4	1253	2	S62544	hypothetical protein	8.62e-02	
41	124	1.3	257	2	B69213	protein-export membra	1.14e-01	
42	124	1.3	286	2	S73424	spermidine/putrescine	1.14e-01	
43	123	1.3	327	2	S48839	guanine nucleotide re	1.14e-01	
44	124	1.3	469	2	G70699	probable roda protein	1.14e-01	
45	124	1.3	614	2	S58306	hypothetical protein	1.14e-01	

## ALIGNMENTS

RESULT ENTRY	1	55:525	#type complete	probable membrane protein YPL006w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES	hypothetical protein YPL11w: hypothetical protein YP8132.07			
ORGANISM	#formal_name Saccharomyces cerevisiae			
DATE	08-May-1995 #sequence_revision 21-Jul-1995 #text_change 06-Feb-1998			
ACCESSIONS	S52525; S59687			
REFERENCE	S52519			
#authors	Badcock, K.; Churcher, C.			
#submission	submitted to the EMBL Data Library, February 1995			
#accession	S52525			
##molecule_type	DNA			
##residues	1-1170 ##label BAD			
##cross-references	EMBL:248483; NID:9683777; PID:9683784; MIPS:YPL006w			
##experimental_source	strain AB972			
REFERENCE	S59677			
#authors	Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Eiseen, J.D.; Storms, R.K.; Vo, D.H.; Wang, Y.; Winnett, E.			
#submission	submitted to the EMBL Data Library, August 1995			
#description	The sequence of Saccharomyces cerevisiae chromosome XVI left arm.			
#accession	S59687			
##molecule_type	DNA			
##residues	1-1170 ##label HAL			
##cross-references	EMBL:033353; NID:965076; PID:965087; MIPS:YPL006w			
GENETICS				
#gene	SGD:NCRL			
##cross-references	SGD:S0005927; MIPS:YPL006w			
#map_position	16L			
KEYWORDS	transmembrane protein			
FEATURE	1-17			
1-17	258-274			
341-357	751-767			
667-683	998-714			
614-630	1004-1020			
1027-1043	1051-1067			
1103-1119	1137-1153			
SUMMARY	#length 1170			

Query Match	2.9%;	Score 271;	DB 2;	Length 1170;
Best Local Similarity	21.2%;	Pred. No. 1.47e-23;		
Matches	36;	Conservative	71;	Mismatches 57; Indels 6; Gaps 5;
Db	557	DISFVATSLMFLYATWALRRDG-KTRLLGISGLIVLIVASIVCAAGFTLFGLSFTL	615	
Oy	305	ELVPTVAFMLVEAVYFSVRKIDVFRSRFLTALCSVITTAGSLAMSLGICFFGFTLSL	364	
Db	616	IIAIVIFPLIAGIDINDIFLTHEYDRNCRQKREYSIDDKIISAGRSPTSLMLCQT	675	
Oy	365	QSKDIFPYLVILVGLSNLVTIKSVSM-DET-FPVKIRVAQALSKGEMHLSKTLTLEI	421	
Db	676	GCFIIAFAVY-MPAVHNFAITSYTSVIFNCVLDLTAYVSLILEYKRSNY	724	
Oy	422	T-IITIGLTFVPYIOEFCIFALYVGLISDPMLOMLFSTILANIKREY	470	
RESULT	2			
ENTRY	T13952	#type complete		
TITLE		membrane protein p1ch2 - mouse		
ORGANISM		#forma_name Mus musculus #common_name house mouse		
DATE		20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999		
ACCESSIONS	T13952			
REFERENCE	217830			
#authors		Motoyama, J.; Takabatake, T.; Takeshima, K.; Hui, C.		
#journal		Nature Genet. (1998) 18:104-106		
#title		P1ch2, a second mouse Patched gene is co-expressed with Sonic hedgehog.		
#cross-references	MUID:98122566			
#accession	T13952			
#status		preliminary; translated from GB/EMBL/DBJ		
#molecule_type	mRNA			
#residues	1-1182	#label MOT		
#cross-references	EMBL:AB010833; NID:d1179188; PID:d1025611; PIDs:BA24651.1			
GENETICS		#experimental_source strain BALB/cCrSlc		
#gene		p1ch2		
KEYWORDS		transmembrane protein		
SUMMARY		#length 1182 #molecular_weight 128585 #checksum 5394		
Query Match	2.4%;	Score 223;	DB 2;	Length 1182;
Best Local Similarity	21.0%;	Pred. No. 9.23e-16;		
Matches	33;	Conservative	65;	Mismatches 57; Indels 2; Gaps 2;
Db	401	GYLLMLAYACVTMLRMDCASQAGVLAGVLLVALAVASGLICALGITFENAAATVOVP	460	
Oy	312	AFMLVFAVYFSVRKIDVFRSRFLTALCSVITTAGSLAMSLGICFFGFTLSLQSKDIFP	371	
Db	461	FLAIGIVDDIFLLAHAFKAPDPTPLPEKMGCLASGTGSAVLSYNNMAVFMAALVP	520	
Oy	372	YLTVIVGLENLSTIKSVSMDETFDKIRVAQALSKGEMHLSKTLTLE-TIITIGLAT	430	
Db	521	-IPALRAFSLQAAIVGSCNPAAVMLVPALISDLRR	556	
Oy	431	FVPYIOEFCIFALYVGLISDPMLOMLFSTILANIKR	467	
RESULT	3			
ENTRY	A31898	#type complete		
TITLE		hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - sea urchin (Strongylocentrotus purpuratus)		
ORGANISM		#forma_name Strongylocentrotus purpuratus #common_name purple urchin		
DATE		10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999		
ACCESSIONS	A31898			
REFERENCE	A31898			
#authors		Woodward, H.D.; Allen, J.M.C.; Lennarz, W.J.		
#journal		J. Biol. Chem. (1985) 263:18411-18418		
#title		3-hydroxy-3-methylglutaryl-coenzyme A reductase of the sea		

```
#cross-references MUID:89054023      urchin embryo. Deduced structure and regulatory properties
#accession A31898                    ##molecule-type mRNA
##residues 1-932 ##label WOO
##cross-references GB:J04200; NID:g161522; PIDN:AAA30060.1; PID:g161523
##note the authors rearranged portions of the coding region in Figure 2, and the above sequence is taken directly from Figure 3: it matches the translation of the nucleotide sequence that the author submitted to Genbank
#####
#note the authors translated the codon GCA for residue 805 as Glu
REFERENCE A28367
#authors Woodward, H.D.; Allen, J.M.C.; Lennarz, W.J.
#journal J. Biol. Chem. (1988) 263:2513-2517
#title 3-hydroxy-3-methylglutaryl coenzyme A reductase in the sea urchin embryo is developmentally regulated.
#cross-references MUID:88115403
#accession A28367
##molecule-type mRNA
##residues 689-735 ##label WO2
COMMENT This transmembrane glycoprotein of the endoplasmic reticulum is involved in the control of cholesterol biosynthesis.
CLASSIFICATION superfamily hydroxymethylglutaryl-CoA reductase (NADPH)
KEYWORDS cholesterol biosynthesis; endoplasmic reticulum; glycoprotein; NADP; oxidoreductase
FEATURE 279,850,886,930 #binding site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 932 #molecular-weight 100965 #checksum 3584
Query Match 2.3%; Score 212; DB 1; Length 932; Best Local Similarity 24.0%; Pred. No. 4.87e-14; Matches 36; Conservative 58; Mismatches 53; Indels 3; Gaps 3;
Db 72 LAVAYVLIQFKKLTGTSKIILGAGLFTFFSFLFSANVHLFGLELT-GLNELAPFL 130
OY 315 LVFAYVFYSVKRIDFVSRLACSVITTAGSIAMSLGICFFGGTLTISQSDIRPYLV 374
Db 131 LLIDLTFASALTRPALSTONEVDVIARGMALLPTITLDIVVTLLV-SIGTMSSIR 189
Oy 375 ILVGLENSIVTKSVSMDETFDVKIKVAQALSKEGNHIS-KLLDLPTITLLTGATFPV 433
Db 190 KMEVPCCGIIISLIANTFVFMTFFPACLST 219
OY 434 VIOEFCIFAIYGLISDFMLQMLLFSTILLAM 463
RESULT 4
ENTRY T05663 #type complete
TITLE hypothetical protein F22I13.120 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
ACCESSIONS T05663
REFERENCE T05420
#authors Bevan, M.; Wedler, H.; Kutner, M.; Wandut, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, February 1999
#accession T05663
##molecule-type DNA
##residues 1-1055 ##label BEV
#cross-references EMBL:AL035319
#experimental_source cultivar Columbia; BAC clone F22I13
GENOTYPE
#map_position 4
#introns 24/3: 60/3: 99/3: 150/1: 193/1: 216/3: 278/3: 297/2: 336/3: 364/3: 396/3: 419/3: 511/3: 548/3: 577/3: 620/3: 667/1: 735/1: 747/3: 775/3: 825/3: 874/1: 895/3: 927/1: 958/3: 961/2: 1003/3: 1026/3
F22I13.120
```

Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
57-78	#domain transmembrane	#status predicted	#label TM2\	
90-114	#domain transmembrane	#status predicted	#label TM3\	
124-149	#domain transmembrane	#status predicted	#label TM4\	
160-187	#domain transmembrane	#status predicted	#label TM5\	
192-220	#domain transmembrane	#status predicted	#label TM6\	
315-339	#domain transmembrane	#status predicted	#label TM7\	
340-887	#domain catalytic hydrophilic	#label HY\		
281	#binding-site carbohydrate (asn)	(covalent)	#status predicted	
SUMMARY	#length 887	#molecular-weight 97080	#checksum 872	
Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
61	DIILITIRCIALLIYFQFONLRQSGKYILGILGFTITSSPFVS-TVYIHF-LDKEL	118		
305	ELVPYVAFMVFAYVSVSRKIDVFRSRLFLALCSVITTAGSLMSGLGCEFFGLITSL	364		
119	TGLNALDFEFLILIDLSASALAKFALNSODEYRENIAROMALIGPTFLIDALVECLV	178		
365	QS-KDIFPFIYVLVLENSLVITKSVSMDETFVDKIRVAQALSREGHHISKTLTLEITI	423		
179	IGVGTMSGVROLEIMCCFCGMSVLANYFVPMTFPACVSLYLELSR	224		
424	LITGLATFVPVYIQEFICFAIVGLISDFMLQMLFST-I-LAMNIKR	467		
RESULT ENTRY TITLE	A23586	#type complete		
ENTRY TITLE	hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) - golden hamster			
ORGANISM	10-Sep-1999	#sequence-revision 10-Sep-1999	#text-change	
DATE	10-Sep-1999			
ACCESSIONS	A23586			
REFERENCE	A23586			
#authors	Skalniak, D.G.; Simoni, R.D.			
#journal	DNA (1985) 4:439-444			
#title	The nucleotide sequence of Syrian hamster HMG-CoA reductase cDNA.			
#cross-references	MU01:86135263			
#accession	A23586			
#molecule	type mRNA			
#residues	1-887	#label SKA		
CLASSIFICATION	#superfamily hydroxymethylglutaryl-CoA reductase (NADPH)			
KEYWORDS	oxidoreductase; transmembrane protein			
SUMMARY	#length 887	#molecular-weight 96970	#checksum 74	
Query Match	2.0%;	Score 183;	DB 1;	Length 887;
Best Local Similarity 18.7%;	Pred. No. 1,20e-09;			
Matches 31;	Conservative 68;	Mismatches 62;	Indels 5;	Gaps 5;
61	DIILITIRCIALLIYFQFONLRQSGKYILGILGFTITSSPFVS-TVYIHF-LDKEL	118		
305	ELVPYVAFMVFAYVSVSRKIDVFRSRLFLALCSVITTAGSLMSGLGCEFFGLITSL	364		
119	TGLNALDFEFLILIDLSASALAKFALNSODEYRENIAROMALIGPTFLIDALVECLV	178		
365	QS-KDIFPFIYVLVLENSLVITKSVSMDETFVDKIRVAQALSREGHHISKTLTLEITI	423		
179	IGVGTMSGVROLEIMCCFCGMSVLANYFVPMTFPACVSLYLELSR	224		
424	LITGLATFVPVYIQEFICFAIVGLISDFMLQMLFST-I-LAMNIKR	467		
RESULT ENTRY TITLE	RDHHE	#type complete		
ENTRY TITLE	hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - human			
ORGANISM	04-Dec-1986	#sequence-revision 04-Dec-1986	#text-change	
DATE	11-Jun-1999			

[illegible]

QY	Db	132	LSKATVLAQFALSRSOD-EYKHNIAKGIMLGPITITLDTVEFLVIGVGMLSGVRLEY	190
QY	73	XYOFCFLKLGSKYTLGAGLDTFVSSVFSSVYINFGSVOV-DIKRALFEFLILD	131	
QY	319	YVYFSVKRIDFNSRFLFLACSVTTTNGSLAMSGLGCLFFEGTLTISQSKDIFPVILVIG	378	
QY	379	LENSLVITK-SVSMDETFPVYKIKVAQALSKEGWHSKTLTLEITLITIGLAFVVIDG	437	
Db	191	LCGRACMSVYVNVVMTFPACSL	216	
QY	438	FCIFAIVGLLEDFMLQMLFLFSTILAM	463	
RESULT	9			
ENTRY	A35728	#type complete		
TITLE	hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - African clawed frog			
ORGANISM	#formal_name Xenopus laevis #common_name African clawed frog			
DATE	10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999			
ACCESSIONS	A35728			
REFERENCE	A35728			
#authors	Chen, H.; Shapiro, D.J.			
#journal	J. Biol. Chem. (1990) 265:4622-4629			
#title	Nucleotide sequence and estrogen induction of Xenopus laevis 3-hydroxy-3-methylglutaryl-coenzyme A reductase.			
#cross-references	MUID:90170974			
#accession	A35728			
#status	preliminary			
#molecule_type	mRNA			
#residues	1-883 ##label CHE			
#cross-references	GB:M29258; NID:q214236; PIDN:AAA49740.1; PID:q214237			
CLASSIFICATION	#superfamily hydroxymethylglutaryl-CoA reductase (NADPH)			
KEYWORDS	NADP; oxidoreductase; transmembrane protein			
SUMMARY	#length 883 #molecular_weight 96719 #checksum 6785			
Query Match	1.9%; Score 179; DB 1; Length 883;			
Best Local Similarity	18.7%; Pred. No. 4,64e-09;			
Matches	31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;			
Db	61	DIITITRCALAIYITFOFONLRQLGSKYTLGAGLDTFSSFVS-TYVHF-LDKEL	118	
QY	305	ELPVYTAFMVFAFYVFSVKRIDFNSRFLFLACSVTTTNGSLAMSGLGCLFFEGTLISL	364	
Db	119	TGLNEALPEFLLIDLSKASALAKFALSSNODERDNIARGAIIQPTTLALVECLV	178	
QY	365	QS-KDIPFYIIVIGLENSLVITKSVSMDETFPVYKIKVAQALSKEGWHSKTLTLEITI	423	
Db	179	IGVGTMSGVROLEIMCCFGCMSVLANTFAFMTEFPACSVLLEISR	224	
QY	424	LTIGLATEFVPVIOGFCIFAIVGLLEDFMLQMLFLFST-I-LAMNIKR	467	
RESULT	10			
ENTRY	S32572	#type complete		
TITLE	hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - fruit fly (Drosophila melanogaster)			
ORGANISM	#formal_name Drosophila melanogaster			
DATE	06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 11-Jun-1999			
ACCESSIONS	S32572			
REFERENCE	S32572			
#authors	Gerltner, F.B.; Chiu, C.Y.; Richter-Mann, L.; Chin, D.J.			
#journal	Mol. Cell. Biol. (1988) 8:2713-2721			
#title	Developmental and metabolic regulation of the Drosophila melanogaster 3-hydroxy-3-methylglutaryl coenzyme A reductase.			
#cross-references	MUID:88302188			
#accession	S32572			
#status	preliminary			
#molecule_type	DNA			
#residues	1-916 ##label GER			



[illegible]

```

                                PIDN:AAA68258.1; CESP:K10B2.1
GENETICS
#experimental_source strain Bristol N2
#gene
#introns
#accession
#description
#status
#molecule_type DNA
#residues
#cross-references
#experimental_source strain Bristol N2
SUMMARY
#length 701 #molecular-weight 80319 #checksum 5749
Query Match
Best Local Similarity 26.0%; Score 155; DB 2; Length 701;
Matches 33; Conservative 40; Mismatches 48; Indels 6; Gaps 6;
Db 371 ITIRVIVGHRRAVNVVDFDRYISASGDRITKYMSMDTLEFVRLTAGHRRGTAQLQYR 430
QY 1005 ITLEGVRLAHQOPITCMQVNDMVFVTSQDHTLKYCLNKSDVEYTLGHGCPYTCFV 1063
Db 431 GR-LV-VSGS-SDNTRIR-LMDIHSGVCLRVLEGHLELYRCIRFDEKRIYSGAYDGKIKVW 486
QY 1064 DRWPGTGGSGSQDGLLCVWDLFTGACMYNIOAHGNAVSLACAPSYVLSLGTDERICVW 1123
Db 487 D-LQAL 492
QY 1124 ERFQGNL 1130

RESULT 14
ENTRY
TITLE
#type fragment
#hypothetical protein 1 (cosmid c29B6) - fission yeast
#(Schizosaccharomyces pombe) (fragment)
#formal_name Schizosaccharomyces pombe
#accession
#status
#molecule_type DNA
#residues
#cross-references EMBL:266525; NID:g1044926; PID:g1044927
GENETICS
#map_position 1
#classification
#superfamily WD repeat homology
FEATURE
18-51
104-137
#domain WD repeat homology #label WD1\
#domain WD repeat homology #label WD2
SUMMARY
#length 267 #checksum 7169
Query Match
Best Local Similarity 26.8%; Score 145; DB 2; Length 267;
Matches 33; Conservative 35; Mismatches 52; Indels 3; Gaps 3;
Db 61 FGHIDNVLGAVVSENYIISSRNDRHARVRLDATSPAECMHVLRGHLASVNVQSSKT 120
QY 1012 LAHQPIPCMQVNDMVFVTSQDHTLKYCLN-KSDVEYTLGHGCPYTCFV-D-RWQP 1068
Db 121 GLIYVASSDRFLRWDTITGHCIRIHAHQGIACAQYNGKFIYSGSSDLTIRIFEASSG 180
QY 1069 GTGGSGSQDGLLCVWDLFTGACMYNIOAHGNAVSLACAPSYVLSLGTDERICVW 1128
Db 181 KLL 183
QY 1129 NL 1131

RESULT 15
ENTRY
TITLE
#type complete
#protein-export membrane protein secf related PAB1176 -
#pyrococcus abyssi (strain Orsay)
#formal_name Pyrococcus abyssi
#accession
#status
#molecule_type DNA
#residues
#cross-references
#experimental_source strain Orsay
SUMMARY
#length 308 #molecular-weight 33806 #checksum 9284
Query Match
Best Local Similarity 25.8%; Score 150; DB 2; Length 308;
Matches 40; Conservative 51; Mismatches 53; Indels 11; Gaps 11;
Db 157 ITYAFIGMA-IVFLFFRIPLPSLTVFSASFSDMVIVAVALMNIIGIELS-QAT-IAALLM 213
QY 317 FAYVYFSVRKIDVRS-RPLL-ALCSVITTAGSLAMSLGCFEFGLTISQSKDIFPYLV 374
Db 214 LIGYSVDSNIIITRLKKE-FIVEEAYSSL-RTGFTWSTTTGLALISLWL-FST-AK 269
QY 375 ILV-GLNSLVITTSVYSMDTEPDVKIRVAQALSKEGWHISKLTLLTITLIGLATFVP 433
Db 270 VIDDIASVLRGLADPMNTWILNAGVLRMYIAKR 304
QY 434 VIQFCIFAIYGLSDPMLQMLFSTILAMNI-KR 467

Search completed: Sun Sep 3 12:45:00 2000
Job time : 106 secs.
```

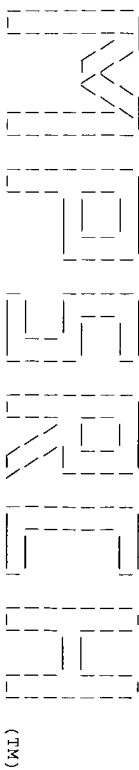
```

                                PIDN:AAA68258.1; CESP:K10B2.1
GENETICS
#experimental_source strain Bristol N2
#gene
#introns
#accession
#description
#status
#molecule_type DNA
#residues
#cross-references
#experimental_source strain Bristol N2
SUMMARY
#length 701 #molecular-weight 80319 #checksum 5749
Query Match
Best Local Similarity 26.0%; Score 155; DB 2; Length 701;
Matches 33; Conservative 40; Mismatches 48; Indels 6; Gaps 6;
Db 371 ITIRVIVGHRRAVNVVDFDRYISASGDRITKYMSMDTLEFVRLTAGHRRGTAQLQYR 430
QY 1005 ITLEGVRLAHQOPITCMQVNDMVFVTSQDHTLKYCLNKSDVEYTLGHGCPYTCFV 1063
Db 431 GR-LV-VSGS-SDNTRIR-LMDIHSGVCLRVLEGHLELYRCIRFDEKRIYSGAYDGKIKVW 486
QY 1064 DRWPGTGGSGSQDGLLCVWDLFTGACMYNIOAHGNAVSLACAPSYVLSLGTDERICVW 1123
Db 487 D-LQAL 492
QY 1124 ERFQGNL 1130

RESULT 14
ENTRY
TITLE
#type fragment
#hypothetical protein 1 (cosmid c29B6) - fission yeast
#(Schizosaccharomyces pombe) (fragment)
#formal_name Schizosaccharomyces pombe
#accession
#status
#molecule_type DNA
#residues
#cross-references EMBL:266525; NID:g1044926; PID:g1044927
GENETICS
#map_position 1
#classification
#superfamily WD repeat homology
FEATURE
18-51
104-137
#domain WD repeat homology #label WD1\
#domain WD repeat homology #label WD2
SUMMARY
#length 267 #checksum 7169
Query Match
Best Local Similarity 26.8%; Score 145; DB 2; Length 267;
Matches 33; Conservative 35; Mismatches 52; Indels 3; Gaps 3;
Db 61 FGHIDNVLGAVVSENYIISSRNDRHARVRLDATSPAECMHVLRGHLASVNVQSSKT 120
QY 1012 LAHQPIPCMQVNDMVFVTSQDHTLKYCLN-KSDVEYTLGHGCPYTCFV-D-RWQP 1068
Db 121 GLIYVASSDRFLRWDTITGHCIRIHAHQGIACAQYNGKFIYSGSSDLTIRIFEASSG 180
QY 1069 GTGGSGSQDGLLCVWDLFTGACMYNIOAHGNAVSLACAPSYVLSLGTDERICVW 1128
Db 181 KLL 183
QY 1129 NL 1131

RESULT 15
ENTRY
TITLE
#type complete
#protein-export membrane protein secf related PAB1176 -
#pyrococcus abyssi (strain Orsay)
#formal_name Pyrococcus abyssi
#accession
#status
#molecule_type DNA
#residues
#cross-references
#experimental_source strain Orsay
SUMMARY
#length 308 #molecular-weight 33806 #checksum 9284
Query Match
Best Local Similarity 25.8%; Score 150; DB 2; Length 308;
Matches 40; Conservative 51; Mismatches 53; Indels 11; Gaps 11;
Db 157 ITYAFIGMA-IVFLFFRIPLPSLTVFSASFSDMVIVAVALMNIIGIELS-QAT-IAALLM 213
QY 317 FAYVYFSVRKIDVRS-RPLL-ALCSVITTAGSLAMSLGCFEFGLTISQSKDIFPYLV 374
Db 214 LIGYSVDSNIIITRLKKE-FIVEEAYSSL-RTGFTWSTTTGLALISLWL-FST-AK 269
QY 375 ILV-GLNSLVITTSVYSMDTEPDVKIRVAQALSKEGWHISKLTLLTITLIGLATFVP 433
Db 270 VIDDIASVLRGLADPMNTWILNAGVLRMYIAKR 304
QY 434 VIQFCIFAIYGLSDPMLQMLFSTILAMNI-KR 467

Search completed: Sun Sep 3 12:45:00 2000
Job time : 106 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:39:57 2000; Maspar time 87.99 Seconds  
974.755 Million cell updates/sec

Tabular output not generated.

Title: >US-09-332-522B-6  
Description: (1-1237) from US09332522B.Pep  
Perfect Score: 9226  
Sequence: 1 MKNKTKSKHFKAMKFWPAC.....CDYGNELRVRFPIVADKCH 1237

Scoring table: PAM 150  
Gap 11

Searched: 225878 seqs, 59334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-phc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 55.496; Variance 103.793; scale 0.535

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1265	13.7	1277	4	Q12770	KIAA0199 PROTEIN (FRAG	2.05e-238
2	1248	13.5	1276	11	P97260	SREBP CLEAVAGE ACTIVAT	1.23e-234
3	565	7.2	1032	5	O18968	D2013.8 PROTEIN.	4.43e-107
4	297	3.2	1086	3	O43043	HYPOTHETICAL 125.1 KD	1.29e-31
5	271	2.9	1170	3	Q12200	HYPOTHETICAL 132.6 KD	1.04e-26
6	248	2.7	608	11	Q922A1	PATCHED (FRAGMENT).	1.84e-22
7	245	2.7	800	5	O9XWL9	Y38F1A.3 PROTEIN.	4.25e-22
8	244	2.6	587	5	O44083	SEL-10.	9.82e-22
9	237	2.6	1278	4	O15118	NIEMANN-PICK C DISEASE	1.81e-20
10	230	2.5	881	5	O16307	T21H3.2 PROTEIN.	3.27e-19
11	229	2.5	1146	4	O95856	PATCHED 2.	4.83e-19
12	229	2.5	1203	4	O9Y163	PATCHED 2.	4.93e-19
13	229	2.5	1203	4	O9Y341	RECEPTOR PROTEIN PATCH	4.93e-19
14	234	2.5	1243	13	O9M5T6	PATCHED-2 PROTEIN.	6.28e-20
15	224	2.4	1278	11	O35604	NPCL.	3.82e-20
16	218	2.4	1318	5	O9XYP5	PUTATIVE HEDGHOG RECE	4.37e-17
17	204	2.2	840	5	O51894	T07H8.6 PROTEIN.	1.19e-14
18	205	2.2	933	5	O51129	SIMILAR TO DROSOPHILA	7.99e-15
19	196	2.1	889	5	P91346	SIMILAR TO DROSOPHILA	2.76e-13
20	191	2.1	936	5	O9XW22	Y18D10A.7 PROTEIN.	1.92e-12

RESULT ID	1	PRELIMINARY:	PRT:	1277 AA.	ALIGNMENTS
AC	Q12770				
AD	Q12770				
DT	01-NOV-1996 (Trembl,rel. 01, Created)				
DT	01-NOV-1996 (Trembl,rel. 01, Last sequence update)				
DT	01-MAY-1999 (Trembl,rel. 10, Last annotation update)				
DE	KIAA0199 PROTEIN (FRAGMENT).				
GN	KIAA0199.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 96281124.				
RA	MGASE.T.; SEKI N.; ISHIKAWA K.; TANAKA A.; NOMURA N.;				
RT	"Prediction of the coding sequences of unidentified human genes. V.				
RT	The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by				
RL	analysis of cDNA clones from human cell line KG-1."				
RL	DNA Res. 3:17-24(1996).				
DR	EMBL: D83782; BAA12111.1.				
DR	PFAM: PF00400; WD40.3.				
FT	NON_TER				
SO	SEQUENCE 1277 AA; 139495 MW; 10EB5099 CRC32;				

Query Match	13.7%; Score 1265; DB 4; Length 1277;
Best local Similarity	34.5%; Pred. No. 2.05e-238;
Matches	277; Conservative 194; Mismatches 282; Indels 71; Gaps 44;
Db	64 PPVDSDRKQEPTEQPEWYGAFA-VYQIYKSSVFPNKLILAVFRSPISRAFQ 122
Qy	81 PQLVNSSTTRSPPLLPMAQSSPAFFVQOITRTSVLPTEGMQMDARRAPLHEVF 140
Db	123 LVEIRNRVLRDSSGIRLELCQVYDLPGLR-KLRNLRLPHNGCLLSGNQWONWE 181
Qy	141 LLEIVRNH--OSSNKRRLIEHVDVVKGTGQDQIDPEEGCLLSPANLMTQNSQ 198
Db	182 RFHADPDITGIHOHEPKTQTS-ATLKDLFGVPGKSGSLVT-R-K-RMVSYTITLV 237
Qy	199 NFDTDITNLTNIIFQY-N-LQSKVSAEMFGPLMDGTGKRRLPARSIIQYALTLF 256
Db	238 FOHYHAKFLGSLRLRLMLAPS-PNCSIRAE---SLVHVFKEEIGVAELIPVTYIIL 293
Qy	257 LKHNDMEYDITLKEKRLRHRYPPLPLASASAEPTIYIFYPGEYRMWELPVYVAFMLV 316

```

Db 294 FAYIFSTRKIDMKVSKWGLAALAAVTVLSSLSMSGVLCITGLPTLNGEIPFYLVVY 353
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 317 FAYVYFSRKIDVFRSRLFLALCVITTAGSLAMSLGICFFGGLTSLQSKDIPFLVLL 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 IGLENVLITKSVSTPVDELKRLRAOGLSSEWSIMKNMATELGIIIGFTLVPAQ 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 377 VGLSNSLVITKSVSMDETFDKIRAOGLSKEGWHISTLTETITLTIGLATEVPVIO 436
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 EECLEAVVGLVSDPFLQMLFETTVLSIDIRRMELADLNKRLPPE--ACLPSSA-----K-- 464
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 437 EECIRAIYGLLSDPMLQMLFSTILAMNKREYTAENKHLKMLLSCGAGRODFRF 496
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 465 ---P-VGQ--PTRYER-Q---LA-VRST-PHTITL-OPSSFRMLRPLKRLVYFLAR 510
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 497 GAAPALPFPVPGTQFORSOSHPRKLCFADPASPVSDDRTSLVNGHSPSEQRIPIKRIIVFMAR 556
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 511 TRLAORLIMAGTVVMGILVYTDPAGLRNYLAAQYTESPFLSEGALAPRVPSSMLPSPH 570
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 557 TRFQRAFMIMIMVIMCSTIY-N-SG--YLEQLFSMQSN-GTMT-ATLELOR-RLQAGR 608
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 PD-PAF-SIFPPDAKLPENOTSPG-ESPERGGPAVVVDSPVEYTWGPEDE-ELMRKL 626
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 609 GAVVSFEFGWQADGQRATSPSGSGSTPIKAPLADINETA-EEMRLRYPSPDLNTFL 667
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 627 SFRHWPFLFSYNTILAKRYISLLPYIPVTLRLNPREALGHRPODG--RS-AMPP-PGP 682
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 668 SNFHMSTIKOYNISLSGHVYVTLPTIRLSHAIAPELATLLRNPOBOLQONTOMKALAA 727
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 683 I-PACHWENG-PKPGCGVQAHGDVTLKYVALGLATGYL--V-LILLC-L---YRVLC 732
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 728 LDPDFNDDDVRESPPVMAEG-LPLVPKSPMEIFPAILLCISTFVLDTYIMVTFRCIC 786
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 723 PRNYGGLGGPGRRRRGELPCDDYGYAPPETEIVPLVNLGHLMDIECLASDMLVSCCL 792
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 767 TRNVEM---RSWHESEMP---KYOTEDILGCVPTQIGHHRIECLVSDAYIISCL 840
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 793 AGHVCVMDAQOTDCLTRIPRGRQ 816
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 841 KQIRVMDARSGEOLTSISRSIQ 864
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 2  
 ID P97260 PRELIMINARY; PRT: 1276 AA.  
 AC P97260;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)  
 DE SREBP CLEAVAGE ACTIVATING PROTEIN.  
 GN SREBP.  
 OS Cricetus griseus (Chinese hamster).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RX MEDLINE: 97053781.  
 RA HUA X., NOTHURFF A., GOLDSTEIN J.L., BROWN M.S.;  
 RT "Sterol resistance in CHO cells traced to point mutation in SREBP  
 cleavage-activating protein.";  
 RL Cell 87:415-426(1996).  
 DR EMBL: U67060; AAB19103.1; -.  
 DR PFM: PF00400; MD40; 3.  
 DR PRINTS: PR00320; GPROTEINBRPT.  
 SQ SEQUENCE 1276 AA; 139512 MW; FBD05617 CRC32;

Query Match 13.5%; Score 1248; DB 11; Length 1276;  
 Best Local Similarity 34.3%; Pred. No. 1.23e-234;  
 Matches 280; Conservative 187; Mismatches 287; Indels 63; Gaps 43;

Db 47 LPLPGGPFVSTPKVDYSP-PPV-DSHKQGEPSQEPMYGAPPA-YIQQILFVSSVS 103  
 Oy 60 LNTSSITTPQEPHPSGEPWPPEOVNLSTTDSRPLPLPMAQSSPAFFYVQDITLTSTVL 119

```

Db 104 PWHKNLAADVFERPLPSRAFOVEIRNHVLDSSGTSLEEDVLOQTDLLPGLR-KIRN 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 120 PWTGEMQMLDAFFRAPLHVEVFKLLEIVRNH--QSENKRTLEHNCILHVDNRKTHGOLDQ 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 LLPHEGCLLSPGNWQMDMERFADPDIIGTHQHEKTLQTS-ATLKDLFCYGPGRYS 221
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 178 IFPEYGLLSPANLWTONSONFTFRDITLNTIFQYH-N-LQSKSVSAEMLPGLPMODT 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 222 GSVLT-R-K-RTVSYTTLVFORYHAKFLSLRLARLMLHPS--PNCISLARN--LVHV 274
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 236 GFKRPLRARSIIQYALTFLKNDMEYLDTLKELLRHPRPLPASASAEPTTIYI 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 HFKEIGIAELIPVTVYIILPAIYFSTRKIDMKVSKWGLAALAAVTVLSSLSMSGVLC 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 296 FYRGEYRMELVPYVAVMLVAVYFVSRIKIDVFRSRLFLALCVITTAGSLAMSLGIC 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 335 TLFGLPTLNGEIPRYLVVYVIGLENVLTKSVSTPVDELKRLRAOGLSSEWSIMK 394
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 366 FFFGLTISLQSKDIPYVIVLIGLNSLVITKSVSMDETFDKIRAOGLSKEGWHISTLC 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 395 NVATELGIIILGYFTLVPAIOEFCIFAVVGLVSDPFLQMLFETTVLSIDIRRMELADLNK 454
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 416 TLTETITLTIGLATFVAVIOEFCIFALVGLSDPMLQMLFSTILAMNKREYTAENK 475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 RLPPE--SCLPSA-----K-----P-VGR--PARYER-Q---LA-VRPA-MPTITL-Q 491
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 476 HLPKMLLCTOGAGQODRFFGGAAPALPFPVGTQFORSOSHPRKLCFADPASPVSDDRTSLVN 535
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 492 PSSFNLPLPRIRYIYVLAFTRLAORLIMAGTVVMGILVYTDPAGLRNYLAAQYTESQ 551
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 536 GHSSQEQRIPIKRIKIVNFMARTRFQRAFMIMIMVIMCSTIY-N-SG--YLEQLFSMQS 590
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 PLGESSLGMVPPSVGLPASRPD-PAF-SIFPPDAKLPENOTVPG-ELPEHAPAEVGH 608
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 591 N-GTMT-ATLELOR-RLQAGRGAVSFEFGWQADGQRATSPSGSGSTPIKAPLADIN 647
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 609 DSRAPETVWGPEDE-ELMRRLSFRHWPFLFSYNTILAKRYISLLPYIPVTLRLNPQAL 667
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 648 ET-AEEMRLRYPSPDLNLYFLSNFHMSTIKOYNISLSGHVYVTLPTIRLSHAIAPELAT 706
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 668 EGRORODG--RS-AW-APPESL-PAGLWENG-PKPGGTOAHGDTLLKVALGALAGIV 721
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 707 LIRNPOEOLQOQFQKALAAALDPLDFNDDDVRESPPVMAEG-LPLVPKSPMEIFPAIL 765
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 722 I--V-LLLCLYRLVCPNNYQPGGAGRRRGELPCDDYGYAPPETEIVPLVNLGHLMD 778
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 766 LOCISIFVLCITMVYFYCICITRANAEWRSSWHESE-APYOTEDILGCVPTQIGHHNR 824
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 779 IECLASDMLVSCCLAGQVCVMDAQOTDCLTRIPR 815
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 825 IECLVSDGAYIISCLKQIRVMDARSGEOLTSISRS 861
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3  
 ID Q18968 PRELIMINARY; PRT: 1032 AA.  
 AC Q18968;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)  
 DE D2013.8 PROTEIN.  
 GN D2013.8.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;  
 CC Rhabditiina; Rhabditiodea; Rhabditiidae; Peloderiinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MORITMORE B.;  
 RL Submitted (Jan-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,







DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)  
DE PATCHED 2.  
GN  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 99135908.  
RA SMYTH I., NARANG M.A., EVANS T., HEIMANN C., NAKAMURA Y.,  
RA CHENEVIX-TRENCH G., PIETSCH T., WICKING C., MAINWRIGHT B.J.;  
RT "Isolation and characterization of human patched 2 (PCH2), a putative  
RT tumour suppressor gene inbasal cell carcinoma and medulloblastoma on  
RT chromosome 1p32." 6:291-297(1999).  
RL Hum. Mol.Genet. 8:291-297(1999).  
RL EMBL; AF087651; AAD23553.1; f  
SQ  
SQ SEQUENCE 1203 AA: 130498 MW: 22682BAB CRC32:

Query Match	2.58;	Score 229;	DB 4;	Length 1203;
Best Local Similarity	22.28;	Pred. No. 4.93e-19;		
Matches	35;	Conservative	66;	Mismatches 53;
			Indels 4;	Gaps 4;

D5 401 GYLLMLAYACVTMLRMDCAOSGSGVGLAVLVALAVASGLGALLGITFENNAATTVLP 460  
:::| | : : | : : | : : | : : | : : | : : |  
QY 312 AFMLVFAYVFSVRKIDVFRSRFLALCSVITTTAGSLANSLGLCEFFGLTISLSQNDIFP 371

Db 461 FLAIGIGVDVDFLLAHAFTEALPGT-PLQERMGECLQRTGTSVVLTSINNAFLMAALY 519  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 372 YLAVLVGLENSLVITKSV-SNDETFDVKIRVAQALSKEBHISKTLLTETI-ILTIGLA 429

D6 520 P-IPALRAFSIQAAIVVGCTFVAVMVLPAILSLDLRR 556  
:  
Q7 430 TFPPIQIEFCFAIVGLSDFMQLMESTILLANNKR 467

RESULT 13  
ID O95341  
PRELIMINARY;  
PRT; 1203 AA

DT	01-MAY-1999 (TREMblrel. 10, Created)
DT	01-MAY-1999 (TREMblrel. 10, last sequence update)
DT	01-MAY-1999 (TREMblrel. 10, last annotation update)

GN PTCH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

CC Eucetidae, Erimacres, Caladaniini, Nomiinae, Nomo-  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 99030620.

RT protein family":  
RA "Characterization of two patched receptors for the vertebrate hedgehog  
RA ROSENTHAL A., DE SAUVAGE F.J.;  
RA CARRENIER D., STONE D.M., BRUSH J., RIAN A., ARMANINI M., FRANKS G.,  
RA

SC SEQUENCE 1203 AA: 130573 MW: 318226764 CP0332:  
KL PROC. NATL. ACAD. SCI. U.S.A. 95:13630-13634 (1998)  
DR EMBL; AF091501; AAC79847.1; -.  
KW Receptor.

Query Match	2.58;	Score 229;	DB 4;	Length 1203;
Best Local Similarity	22.28;	Pred. No. 4.93e-19;		
Matches	35;	Conservative	66;	Mismatches 53;
				Indels 4;
				Gaps 4;

Dd 401 GYLLMLAYACVTMLRMDCAOSGSGVGLAGVLLVALAVASGLCALIGTFENAT<sup>Q</sup>VLP 460  
:::|::| :: | :: : :: : :: : | :: : |||| ::|::: :: :::  
Qv 312 AFMVFAYVFSVRKIDVFRSEFLALCSVITTAGSIAMSLGCEFFGLTISLOSDFP 371

D6  
Db

461 FLAGIGVDVDFLLAHFTEALPGT-PLQERMGECLQTGTSVLTSINNAAFILMALY 519

: : : : : | : : : : | : : : : | : : : : |  
0v 372 YLVIVGLNSLYITSKV-SMDETDPKIRVOALSKEEMHISKULLTTET-IITGLIA 4299

Db      520 P-IPALRAFSLOAIVGCTFVAVMVFPAISLDLR 556  
      :::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 430 TFVPVIOEFCIFAIVGLSDFMLQMLLSTILAMNIKR 467

RESULT	14	
ID	Q9W6T6	PRELIMINARY;
		PRT; 1243 AA

DT	01-NOV-1999	(TREMblrel. 12, Created)
DT	01-NOV-1999	(TREMblrel. 12, last sequence update)
DT	01-NOV-1999	(TREMblrel. 12, last annotation update)

GN PTCL.  
OS Brachydanio rerio (zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii

OC Cyprinoidae; Cyprinidae; Rasbora; Danio.  
RN [1]  
RP SEQUENCE FROM N.A.

RT "Characterisation of the second ptc gene in zebrafish."  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases  
 DR EMBL; AJ007742; CAB39726.1; -.

Query Match	2.58;	Score 234;	DB 13;	Length 1243;
Best Local Similarity	22.2%;	Pred. No. 6.28e-20;		

Db 428 VIRIAGYLIMLAACITMLRWDCAKSQAGVIGIIVLTLSVAAGLGICSLGISFNAA 487

Db 488 TTQVLPFLALGVVDVFLLAHAFSEHQNRIPEDRTGCEKLRGTASVLTSSINVT 547

500 GNDIFFILVGLNSDYIN SVSMDEIFDKALNRYZDUSDSMHII  
 548 FFMALIP-IPALRAFSLQAAVVVFENFAMVLLIFPAISMDLYRRE 5933

**RESULT**    15

ID	035604	PRELIMINARY,	FRI,	12/8 AM.
DC	035604;			
DT	01-JAN-1998	(TREMBlrel, 05, Created)		
DT	01-JAN-1998	(TREMBlrel, 05, Last sequence update)		

DT 01-NOV-1999 (IEMBLEI. 12, last annotation update)  
 DE NPCL.  
 GN NPCL.  
 OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 RN [1]  
 PR SEQUENCE FROM N A

RA MEDLINE; 97362324.  
RA LOFTUS S.K., MORRIS J.A., CARSTEA E.D., GU J.Z., CUMMINGS C.,  
RA BROWN A., ELLISON J., OHNO K., ROSENFELD M.A., TAGLE D.A.,  
RA BREWSTER D.C., DRYAN E. T.

RT "Murine model of Niemann-Pick C disease: mutation in a cholesterol  
RT homeostasis gene."; [Science 277:232-235\(1997\)](#).  
RL [PubMed 9100000](#) [PubMed 9100000](#)

DR MGD; MGI:1097712; Npc1.  
SQ SEQUENCE 1278 AA; 142887 MW; 072DA88B CRC32.

Best Local Matches	Similarity	Pred. No.	Conservative	Mismatches	Indels	Gaps
45	26.3%	3.82e-18	61	54	11	9

Qy 306 LVPITVAFMLV-FAIYF-SVRKIDVRSRFLALCSVITTAGSLAMSLGLCEFFGLTIS 363



OY   |   ::|::|: ||::|   :::   :   | |   :   : |   ::|   |:   :  
364   LQSKDIEPPYIWLVLVGLNSLVTITKSVSMDETFDKIRVAQALSKE-GWHISKTL-LT-- 419  
Db    742   SETSAFFEGALSMSPAVHTFSLFAGMAVLIDFLLOITCFVSLGLDKROE 792  
      |   :   : |   :   : |   :   : |   :   : |   :   : |   :   : |   :  
OY    420   -ETITLITGLATFVPVIOEFCIFALVGLSDPMLQMLLFTIIMNKRTE 469

Search completed: Sun Sep 3 12:42:57 2000  
Job time : 180 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
NWSENF  
(TW)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch.pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:38:30 2000; Maspar time 37.66 Seconds

Tabular output not generated. 1017.918 Million cell updates/sec

Title: >US-09-332-522B-6  
Description: (1-1237) from US09332522B.pap  
Perfect Score: 9226  
Sequence: 1 MNKTKSKHFKAWKFWPAC.....CDYGNIRVVRPPIVADKCH 1237

Scoring table: PAM 150  
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 56.799; Variance 99.679; scale 0.570

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	246	2.7	1220	1	PTC1_BRARE	1.08e-23
2	244	2.6	579	1	SEL10_PROTEIN.	2.61e-23
3	240	2.6	1434	1	PTC1_MOUSE	1.52e-22
4	239	2.6	1442	1	PTC1_CHICK	2.37e-22
5	240	2.6	1447	1	PTC1_HUMAN	1.52e-22
6	223	2.4	1182	1	PTC2_MOUSE	2.51e-19
7	212	2.3	932	1	HMDH_STRPU	2.75e-17
8	182	2.0	703	1	WD-REPEAT PROTEIN POP2	6.46e-12
9	180	2.0	732	1	KHMB_DICD1	1.43e-11
10	183	2.0	887	1	HMDH_MESLA	4.33e-12
11	183	2.0	887	1	HMDH_CRIGR	4.33e-12
12	183	2.0	887	1	HMDH_RAT	4.33e-12
13	183	2.0	888	1	HMDH_RABIT	4.33e-12
14	183	2.0	888	1	HMDH_HUMAN	4.33e-12
15	178	1.9	856	1	HMDH_BLAG	2.13e-11
16	179	1.9	883	1	HMDH_XENLA	2.13e-11
17	177	1.9	916	1	HMDH_DROME	4.71e-11
18	177	1.9	1053	1	HMDH_SCHPO	4.71e-11
19	162	1.8	332	1	GLP1_DICD1	1.57e-08
20	152	1.8	605	1	YD15_SCHPO	1.57e-08
21	155	1.7	701	1	YSL1_CAEEL	2.17e-07
22	145	1.6	267	1	YAF1_SCHPO	8.31e-06
23	147	1.6	479	1	PR12_ARATH	4.05e-06

ID	PTC1_BRARE	STANDARD:	PRT:	1220 AA.
AC	Q98864;			
DT	15-FEB-2000 (Rel. 39, Created)			
DT	15-FEB-2000 (Rel. 39, Last sequence update)			
DE	15-FEB-2000 (Rel. 39, Last annotation update)			
GN	PATCHED PROTEIN HOMOLOG 1 (PATCHED 1) (PTC1).			
OS	Brachydanio rerio (zebrafish) (Zebra danio).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;			
OC	Cypriniformes; Cyprinidae; Cyprinidae; Rasbora; Danio.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
KC	TISSUE=EMBRYO;			
RX	MEDLINE: 96379744.			
RA	Concordet J.-P., Lewis K.E., Moore J.W., Goodrich L.V., Johnson R.L.,			
RA	Scott M.P., Ingham P.W.;			
RT	"Spatial regulation of the zebrafish patched homologue reflects the			
RT	roles of sonic hedgehog and protein kinase A in neural tube and somite			
RT	patterning."			
RL	Development 122:2835-2846(1996).			
CC	-1- FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN			
CC	HEDGEHOG (IHH) AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE			
CC	SMOOTHENED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEIN			
CC	SIGNAL (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: DETECTED IN EMBRYONIC PRESOMITIC MESODERM,			
CC	NEUROECTODERM, TISSUE SURROUNDING THE NOTOCHORD, VENTRAL NEURAL			
CC	TUBE.			
CC	-1- DEVELOPMENTAL STAGE: AT ALL STAGES, EXPRESSION CORRESPONDS TO THE			
CC	LOCALIZATION OF SHH. FIRST DETECTED DURING GASTRULATION. BY 36			
CC	HOURS, PTC1 APPEARS IN THE FIRST BRANCHIAL ARCH AND THE POSTERIOR			
CC	MESONEPHROS OF THE FIN BUD; BY 48 HOURS, IN THE HINDBRAIN AND			
CC	FOREBUD.			
CC	-1- INDUCTION: ACTIVATED BY SONIC HEDGEHOG.			
CC	-1- PTM: GLYCOSYLATION IS NECESSARY FOR SHH BINDING (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE PATCHED FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			

24 149 1.6 518 1 TROB\_XENLA BETA-TRCP (BETA-TRANSD 1.97e-06

25 149 1.6 976 1 HMDH\_GIBRU 3-HYDROXY-3-METHYLGLUT 1.97e-06

26 146 1.6 1146 1 KMAA\_DICD1 MYOSIN HEAVY CHAIN KIN 5.91e-06

27 139 1.5 395 1 YLL1\_CAEEL HYPOTHETICAL 43.1 KDA 6.91e-05

28 136 1.5 425 1 SECY\_ADOEI PREPROTEIN TRANSLOCASE 1.95e-04

29 135 1.5 684 1 CC4\_CANAL CELL DIVISION CONTROL 2.75e-04

30 139 1.5 775 1 POP1\_SCHPO WD-REPEAT PROTEIN POP1 6.91e-05

31 137 1.5 779 1 CC4\_YEAST CELL DIVISION CONTROL 1.38e-04

32 138 1.5 1356 1 HET1\_PODAN VEGETABLE INCOMPATIB 9.79e-05

33 129 1.4 282 1 VC53\_METLA HYPOTHETICAL PROTEIN M 2.08e-03

34 128 1.4 318 1 GLP1\_CHLRE GUANINE NUCLEOTIDE-BIN 2.90e-03

35 128 1.4 325 1 GLP1\_SOYBN GUANINE NUCLEOTIDE-BIN 2.90e-03

36 126 1.4 325 1 GLP1\_MEDSA GUANINE NUCLEOTIDE-BIN 5.59e-03

37 126 1.4 334 1 GLP1\_ORYSA GUANINE NUCLEOTIDE-BIN 5.59e-03

38 131 1.4 486 1 PRL1\_ARATH PPL/PP2A PHOSPHATASES 1.07e-03

39 125 1.4 904 1 COPT\_RAT COATOMER BETA' SUBUNIT 7.74e-03

40 131 1.4 905 1 COPT\_HUMAN COATOMER BETA' SUBUNIT 1.07e-03

41 131 1.4 905 1 COPT\_BOVIN COATOMER BETA' SUBUNIT 1.07e-03

42 129 1.4 917 1 FAN\_HUMAN PROTEIN FAN (FACTOR AS 2.08e-03

43 125 1.4 952 1 PMA4\_NICPL PLASMA MEMBRANE ATPASE 7.74e-03

44 125 1.4 1253 1 YAGD\_SCHPO HYPOTHETICAL 143.3 KDA 7.74e-03

45 124 1.3 678 1 SCOB\_EMENT SULFUR METABOLITE REPR 1.07e-02

```

DR EMBL: X98883: CAAG7386.1: -.
DR ZFIN: ZDB-GENE-980526-44: PTC1.
KW Receptor; Transmembrane; glycoprotein.
FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 85 105 POTENTIAL.
FT DOMAIN 106 419 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 420 440 POTENTIAL.
FT DOMAIN 441 449 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 450 470 POTENTIAL.
FT DOMAIN 471 484 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 485 505 POTENTIAL.
FT DOMAIN 506 528 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 529 549 POTENTIAL.
FT DOMAIN 550 558 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 559 579 POTENTIAL.
FT DOMAIN 580 739 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 740 760 POTENTIAL.
FT DOMAIN 761 1016 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1017 1037 POTENTIAL.
FT DOMAIN 1038 1044 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 1045 1065 POTENTIAL.
FT DOMAIN 1066 1072 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1073 1093 POTENTIAL.
FT DOMAIN 1094 1110 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 1111 1131 POTENTIAL.
FT DOMAIN 1132 1143 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1144 1164 POTENTIAL.
FT DOMAIN 1165 1220 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 397 397 POTENTIAL.
FT CARBOHYD 865 865 POTENTIAL.
FT CARBOHYD 888 888 POTENTIAL.
SQ SEQUENCE 1220 AA; 135544 MW; D10A9D04115F532D CRC64;

Query Match 2.7%; Score 246; DB 1; Length 1220;
Best Local Similarity 23.9%; Pred. No. 1, 08e-23;
Matches 38; Conservative 61; Mismatches 58; Indels 2; Gaps 2.

Db 428 GYLMLAVACYTMRKWCARSGAGVAGLVLLVAASVAGLGICSLGSEFNATTVLP 487
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 312 AFMLVFAYVYSVRKIDYFRSRFLALCSVTTTGAISLMSGLCFPGTLTSLQSDIFP 371
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 488 SLAIGIGVDMDFLLGHSTETRSNPFRERFGDCLRGTGVALTGVNNIAFMALVP 547
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 372 YLVIVGVENSLVITIKSVSMDFTFDVKIRVQAQASKGWHITKTLTE-ITLITGLAT 430
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 548 -IPALRAFSLQAAVVVVFNFAMALLIPPAISLDLHRE 585
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 431 FVPVIOEFCIFAIVGLSLDFMLQMLFSTLIAMNKKRE 469
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 2
ID SE10-CAEEL STANDARD; PRT; 579 AA.
AC Q93794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SEL-10 PROTEIN.
GN SEL-10 OR F55B12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Sims M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RC STRAIN-BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
-----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).

CC -----  
CC EMBL: Z79757; CAB02129.1; -  
CC WORKPER; F55212.3; CE16120.  
CC PFAM; PF00646; F-box; 1.  
CC DR PFAM; PF00400; WD40; 7.  
CC DR PRINTS; PR00320; GPROTEINRPT.  
CC DR PROSITE; PS00678; WD\_REPEATS; 5.  
CC KW Repeat; WD repeat.  
CC FT REPEAT 245 274 WD1.  
CC FT REPEAT 286 316 WD2.  
CC FT REPEAT 328 356 WD3.  
CC FT REPEAT 368 396 WD4.  
CC FT REPEAT 408 438 WD5.  
CC FT REPEAT 453 481 WD6.  
CC FT REPEAT 493 522 WD7.  
CC SO SEQUENCE 579 AA; 64275 MW; 0647245AF5964663 CRC64;

Query Match 2.6%; Score 244; DB 1; Length 579;  
Best Local Similarity 38.1%; Pred. No. 2,61e-23;  
Matches 48; Conservative 31; Mismatches 41; Indels 6; Gaps 5;

Db 245 GHEDHVTICQAIHDDVLVTGSDNTLTKWCLDKGEVMTLVGHGHWTSQ-IS--QCGR 301  
QY 1013 AHQ--OPTICMOVVMDWMTFVSGSDHLLKYLCKNSDVEETLHGCGPV-TCLFVRWPGGT 1070  
Db 302 YIVSGSTRTFKVAVSTVSGSLHTLQGHTSVRCAMAGSILVLTGSRPDTLRVWDVESGR 361  
QY 1071 -GGSSSGDGLLCVMDLFTFGACMYNIQAIDGAVSCLACAPSVISLGTDERICVWERFQGN 1129  
Db 362 HLAHLH 367  
QY 1130 LLTTIN 1135

RESULT 3  
ID PTCL\_MOUSE STANDARD; PRT; 1434 AA.  
AC Q61115.  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PATCHED PROTEIN HOMOLOG 1 (PTCL) (PTC).  
GN PCH.  
OS Mus musculus (Mouse).  
OC Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-EMBRYO;  
RX MEDLINE; 96176226.  
RA Goodrich L.V., Johnson R.L., Milenkovic L., McMahon J.A., Scott M.P.;  
RT "Conservation of the hedgehog/patched signaling pathway from flies to  
RT mice: induction of a mouse patched gene by Hedgehog.";  
RL Genes Dev. 10:301-312(1996).  
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN  
CC HEDGEHOG (IHH) AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE  
CC SMOOTHED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEINS  
CC SIGNAL. SEEMS TO HAVE A TUMOR SUPPRESSOR FUNCTION, AS INACTIVATION  
CC OF THIS PROTEIN IS PROBABLY A NECESSARY, IF NOT SUFFICIENT STEP  
CC FOR TUMORIGENESIS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVEL AT 7 DAYS POST COITUM  
CC (DPC), HIGHER LEVEL AT 11 AND 15 DPC, AND PERSISTS AT MODERATE  
CC LEVELS AT 17 DPC. IN THE DEVELOPING EMBRYO, PTC IS FIRST DETECTED  
CC WITHIN THE VENTRAL NEURAL TUBE AND LATER IN THE SOMITES AND LIMB  
CC BUDS. EXPRESSION IN THE LIMB BUDS IS RESTRICTED TO THE POSTERIOR  
CC ECTODERM SURROUNDING THE ZONE OF POLARIZING ACTIVITY. IN THE

Dt		15-FEB-2000 (Rel. 39, Last annotation update)
De	PATCHED PROTEIN HOMOLOG 1 (PTC1) (PTC).	
Gn	PTCH OR PTC.	
Og	Gallus gallus (Chicken).	
Cc	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Oc	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
Oc	Gallus.	
Rn	(1)	
Rp	SEQUENCE FROM N.A.	
Rc	TISSUE=LIMB BUD:	
Rx	MEDLINE: 96205046.	
Ra	Marigo V., Scott M.P., Johnson R.L., Goodrich L.V., Tabin C.J.;	
Rt	"Conservation in hedgehog signaling: induction of a chicken Patchd homolog by Sonic hedgehog in the developing limb.";	
Rl	Development 127:1225-1233(1996).	
Rm	[2]	
Rp	CHARACTERIZATION.	
Rx	MEDLINE: 97064175.	
Ra	Marigo V., Davey R.A., Zuo Y., Cunningham J.M., Tabin C.J.;	
Rt	"Biochemical evidence that patched is the Hedgehog receptor.";	
Rl	Nature 384:176-179(1996).	
Cc	-I FUNCTION: ACTS AS A RECEPTOR FOR SONIC HEDGEHOG (SHH), INDIAN HEDGEHOG (IH), AND DESERT HEDGEHOG (DHH). ASSOCIATES WITH THE SMOOTHENED PROTEIN (SMO) TO TRANSDUCE THE HEDGEHOG'S PROTEINS SIGNAL.	
Cc	-I SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.	
Cc	-I TISSUE SPECIFICITY: EXPRESSION IS SEEN IN THE EMBRYONIC NEURAL TUBE, SCEROTOME, VISCERAL MESODERM, AND LIMB BUD.	
Cc	-I DEVELOPMENTAL STAGE: IN STAGE 10 EMBRIO, EXPRESSION IS SEEN IN NEURAL TUBE, AND AT LOWER LEVELS IN THE NOTOCHORD, EPITHELIAL SOMITES, ENDODERM AND SPLANCHNIC MESODERM. AT STAGE 18, PTC IS BROADLY EXPRESSED IN THE NEURAL TUBE BUT EXCLUDED FROM THE CELLS OF THE FLOOR PLATE. AT STAGE 32, EXPRESSION OCCURS IN THE MESEDERMAL CELLS OF THE GASTROINTESTINAL TRACT.	
Cc	-I INDUCTION: ACTIVATED BY HEDGEHOG; REPRESSED BY ITSELF (PROBABLE).	
Cc	-I PTM: GLYCOSYLATION IS NECESSARY FOR SHH BINDING.	
Cc	-I SIMILARITY: BELONGS TO THE PATCHED FAMILY.	
Cc	-----	
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).	
Cc	-----	
Dr	EMBL: U40074; AAC59898.1; -	
Kw	Receptor; Transmembrane; Glycoprotein.	
Ft	DOMAIN . 101	
Ft	TRANSMEM 102 122 POTENTIAL.	
Ft	DOMAIN 123 436 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 437 457 POTENTIAL.	
Ft	DOMAIN 458 472 CYTOPLASMIC (POTENTIAL).	
Ft	DOMAIN 473 493 POTENTIAL.	
Ft	DOMAIN 494 501 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 502 522 POTENTIAL.	
Ft	DOMAIN 523 547 CYTOPLASMIC (POTENTIAL).	
Ft	TRANSMEM 548 568 POTENTIAL.	
Ft	DOMAIN 569 577 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 578 598 POTENTIAL.	
Ft	DOMAIN 599 747 CYTOPLASMIC (POTENTIAL).	
Ft	TRANSMEM 748 768 POTENTIAL.	
Ft	DOMAIN 769 1026 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 1027 1047 POTENTIAL.	
Ft	DOMAIN 1048 1053 CYTOPLASMIC (POTENTIAL).	
Ft	TRANSMEM 1054 1074 POTENTIAL.	
Ft	DOMAIN 1075 1082 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 1083 1101 POTENTIAL.	
Ft	DOMAIN 1102 1120 CYTOPLASMIC (POTENTIAL).	
Ft	TRANSMEM 1121 1141 POTENTIAL.	
Ft	DOMAIN 1142 1153 EXTRACELLULAR (POTENTIAL).	
Ft	TRANSMEM 1154 1174 POTENTIAL.	
Ft	DOMAIN 1175 1442 CYTOPLASMIC (POTENTIAL).	
Ft	DOMAIN 1442 1442 CYTOPLASMIC (POTENTIAL).	









```

RC STRAIN=AX3;
RA MEDLINE: 97277316.
RA Klancy C.E., Mendoza M.G., Naismith T.V., Kolman M.F., Egejhoft T.T.;
RT "Identification of a protein kinase from Dictyostelium with homology
RT to the novel catalytic domain of myosin heavy chain kinase A.";
RL J. Biol. Chem. 272:11812-11815(1997).
CC -1- FUNCTION: PHOSPHORYLATES THREONINE IN THE C-TERMINAL TAIL REGION
CC OF MYOSIN II HEAVY CHAIN. THIS PHOSPHORYLATION IS CRITICAL IN
CC REGULATING THE ASSEMBLY AND DISASSEMBLY OF MYOSIN II FILAMENT.
CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN HEAVY-CHAIN] = ADP +
CC [MYOSIN HEAVY-CHAIN] PHOSPHATE.
CC -1- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN, A CENTRAL NONREPETITIVE
CC CATALYTIC DOMAIN, AND A C-TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE MCKC / EF-2 PROTEIN KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 090946; AAB50136.1; -
CC DICTYDB: DD01087; MKKB.
CC PFAM: PF00400; WD40; 7.
CC PRINTS: PR00320; GPROTEINBRPT.
CC DR PROSITE: PS00678; WD_REPEATS.5.
CC DR Transferase: Serine/threonine-protein kinase; ATP-binding; Repeat;
CC KM WD repeat.
CC FT NP_BIND 298 303 ATP (POTENTIAL).
CC FT DOMAIN 355 381 POLY-ASN.
CC FT REPEAT 458 485 WD1.
CC FT REPEAT 500 528 WD2.
CC FT REPEAT 540 568 WD3.
CC FT REPEAT 580 608 WD4.
CC FT REPEAT 620 648 WD5.
CC FT REPEAT 660 688 WD6.
CC FT REPEAT 700 730 WD7.
CC FT REPEAT 732 AA; 83167 MM; A7233CABD56D4088 CRC64;
CC SO SEQUENCE

Query Match 2.0%: Score 180; DB 1; Length 732;
Best local similarity 35.7%: Pred. No. 1.43e-11;
Matches 41; Conservative 30; Mismatches 35; Indels 9; Gaps 8;

DB 500 GHEGPVESITQYNOQLFSGSSDSHSIKYWDLKKRLCITFLTEGHDKPVTYVLNDKYL--FS 557
QY 1013 AHQPIICMNVNDVMTGSDHTLKYCLNKSDVEYTLGHGCPV--TCLFVDRMPGTG 1071
DB 558 GS-SDKTIK-VWDLKTLECKYTLFESHAAVKTL-CISGOYLFSGSNDKTIKWD 608
QY 1072 GSGSDGLLCVWDLFTGACMNTIOAHDAVSLACAPS-YVISLGTDER-ICWE 1124

RESULT 10
ID HMDH.MESAU STANDARD: PRT: 887 AA.
AC P09610:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
DE REDUCTASE).
CN HMGCR.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86135263.
RA Skalniak D.G., Simoni R.D.;
RT "The nucleotide sequence of Syrian hamster HMG-CoA reductase cDNA.";

```

---

```

RL DNA 4:439-444(1985).
RN [2]
RN MUTAGENESIS OF HISTIDINE RESIDUES.
RX MEDLINE: 93231987.
RA Darnay B.G., Rodwell V.W.;
RT "His865 is the catalytically important histidyl residue of Syrian
RT hamster 3-hydroxy-3-methylglutaryl-CoA reductase.";
RL J. Biol. Chem. 268:8429-8435(1993).
RN [3]
RN MUTAGENESIS.
RX MEDLINE: 94117431.
RA Ffimpong K., Rodwell V.W.;
RT "The active site of hamster 3-hydroxy-3-methylglutaryl-CoA reductase
RT resides at the subunit interface and incorporates catalytically
RT essential acid:c residues from separate polypeptides.";
RL J. Biol. Chem. 269:1217-1221(1994)
CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE
CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING
CC ENZYME OF STEROL BIOSYNTHESIS.
CC -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
CC -1- ENZYME REGULATION: THE ACTIVITY OF HMG-COA-REDUCTASE IS
CC SUPPRESSED BY EXOGENOUS MEVALONATE.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M12705; AAA37077.1; -
CC PIR: A23586; A23586.
CC DR PFAM: PF00368; HMG-COA_red; 1.
CC DR PRINTS: PR00071; HMGCOARDTASE.
CC DR PROSITE: PS00066; HMG-COA-REDUCTASE_1; 1.
CC DR PROSITE: PS00318; HMG-COA-REDUCTASE_2; 1.
CC DR PROSITE: PS00318; HMG-COA-REDUCTASE_2; 1.
CC DR PROSITE: PS01142; HMG-COA-REDUCTASE_3; 1.
CC DR PROSITE: PS50065; HMG-COA-REDUCTASE_4; 1.
CC KM Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
CC Cholesterol biosynthesis; NADP.
CC FT DOMAIN 1 339
CC FT DOMAIN 340 449 LINKER.
CC FT DOMAIN 450 887 CATALYTIC.
CC FT TRANSSEM 10 39 POTENTIAL.
CC FT TRANSSEM 57 78 POTENTIAL.
CC FT TRANSSEM 90 114 POTENTIAL.
CC FT TRANSSEM 124 149 POTENTIAL.
CC FT TRANSSEM 160 187 POTENTIAL.
CC FT TRANSSEM 192 220 POTENTIAL.
CC FT TRANSSEM 315 339 POTENTIAL.
CC FT ACT_SITE 558 558
CC FT ACT_SITE 766 766
CC FT ACT_SITE 865 865
CC FT CARBOHYD 281 281
CC FT CARBOHYD 517 517
CC FT CARBOHYD 859 869
CC FT MUTAGEN 474 474 H->Q: NO LOSS OF ACTIVITY.
CC FT MUTAGEN 487 487 H->Q: NO LOSS OF ACTIVITY.
CC FT MUTAGEN 558 558 E->D,Q: LOSS OF ACTIVITY.
CC FT MUTAGEN 751 751 H->Q: NO LOSS OF ACTIVITY.
CC FT MUTAGEN 766 766 D->N: LOSS OF ACTIVITY.
CC FT MUTAGEN 860 860 H->Q: NO LOSS OF ACTIVITY.
CC FT MUTAGEN 865 865 H->K,Q: LOSS OF ACTIVITY.
CC FT MUTAGEN 866 866 H->Y: NO LOSS OF ACTIVITY.
CC FT SEQUENCE 887 AA; 96955 MM; BC6F9D1F8CAD5EA5 CRC64;

```

DR	EMBL	L00183	AAA36989.1	-	JOINED.
DR	EMBL	L00166	AAA36989.1	-	JOINED.
DR	EMBL	L00169	AAA36989.1	-	JOINED.
DR	EMBL	L00170	AAA36989.1	-	JOINED.
DR	EMBL	L00171	AAA36989.1	-	JOINED.
DR	EMBL	L00173	AAA36989.1	-	JOINED.
DR	EMBL	L00176	AAA36989.1	-	JOINED.
DR	EMBL	L00177	AAA36989.1	-	JOINED.
DR	EMBL	L00178	AAA36989.1	-	JOINED.
DR	EMBL	L00179	AAA36989.1	-	JOINED.
DR	EMBL	L00180	AAA36989.1	-	JOINED.
DR	EMBL	L00181	AAA36989.1	-	JOINED.
DR	EMBL	L00182	AAA36989.1	-	JOINED.
DR	EMBL	X00494	CAA25189.1	-	JOINED.
DR	PIR	A00357	RDHYE	-	JOINED.
DR	PRIM	PF00368	HMG_COA_red	1.	JOINED.
DR	PRINTS	PR00071	HMGCOADPTASE	-	JOINED.
DR	PROSITE	PS00066	HMG_COA_REDUCTASE_1	1.	JOINED.
DR	PROSITE	PS00318	HMG_COA_REDUCTASE_2	1.	JOINED.
DR	PROSITE	PS01192	HMG_COA_REDUCTASE_3	1.	JOINED.
DR	PROSITE	PS50065	HMG_COA_REDUCTASE_4	1.	JOINED.
KW	Oxidoreductase, Glycoprotein, Endoplasmic reticulum, Transmembrane;				
KW	Cholesterol biosynthesis; NADP				
FT	DOMAIN	1	339	MEMBRANE-BOUND.	
FT	DOMAIN	340	449	LINKER.	
FT	DOMAIN	450	887	CATALYTIC.	
FT	TRANSMEM	10	39		
FT	TRANSMEM	57	78		
FT	TRANSMEM	90	114		
FT	TRANSMEM	124	149		
FT	TRANSMEM	160	187		
FT	TRANSMEM	192	220		
FT	TRANSMEM	315	339		
FT	ACT_SITE	558	558	BY SIMILARITY.	
FT	ACT_SITE	766	766	BY SIMILARITY.	
FT	ACT_SITE	865	865	GENERAL BASE (BY SIMILARITY).	
FT	CARBOHYD	281	281		
FT	CARBOHYD	517	517		
FT	CARBOHYD	869	869		
FT	VARIANT	733	733		
SO	SEQUENCE	887 AA;	97080 MW;	4331E53ADA250E6A CRC64;	
N -> C.					
Query Match	2.0%	Score 183;	DR 1;	Length 887;	
Best Local	Similarity 18.7%;	Pred. No. 4.33e-12;			
Matches	31;	Conservative	68;	Mismatches 62;	Indels 5; Gaps 5.
Db	61	DIILITTCRCALAIYIYFQFONLRQLGSKYILGIAGLFTIFSSVFES-TVVIMF-LDKEL	118		
Qy	305	ELVYITAAFMVFAVYVYFVSRKIDIVFGRFLALCSVITTAGSLAMSLGICFFPGGLISL	364		
Db	119	TGLNEALPEFLLIDLSRASALAKFALSSNDSODEVRENIRGMAIILPPTLLDAVECLV	178		
Qy	365	QS-KDIPPIYVILVGLGNSLVITTSVYSMDPEPVKIRVAQALSKESGMHISKTLLEITI	423		
Db	179	IGVGMGSGVROLEIMCEGCGSVLANFVFETFPACVSLVELSR	224		
Qy	424	LITGLATFVPYIOEFCIFALVGLSLDPMLOMLFST-I-LAMNIKR	467		
RESULT	12				
ID	HMDH_RAT	STANDARD:	PRT:	887 AA.	
AC	P51639:				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	3-HYDROXY-3-METHYLGLOUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA REDUCTASE).				
GN	HMGCR.				
OS	Rattus norvegicus (Rat).				
CC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
PP	[1]				
PP	SEQUENCE FROM N.A.				

RA Hangjong C., Williams D., Shapiro D.;  
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RC TISSUE=LIVER.  
 RX MEDLINE: 85261452.  
 RA Edwards P.A., Kemper E.S., Ian S.-F., Erickson S.K.;  
 RT "Functional size of rat hepatic 3-hydroxy-3-methylglutaryl coenzyme A  
 RT reductase as determined by radiation inactivation.";  
 RL J. Biol. Chem. 260:10278-10282(1985).  
 RN [3]  
 RP SEQUENCE OF 423-443 AND 667-678, AND PHOSPHORYLATION.  
 RC TISSUE=LIVER.  
 RX MEDLINE: 90316098.  
 RA Clarke P.R., Hardie D.G.;  
 RT "Regulation of HMG-CoA reductase: identification of the site  
 RT phosphorylated by the AMP-activated protein kinase in vitro and in  
 RT intact rat liver.";  
 RL EMBO J. 9:2439-2446(1990).  
 RN [4]  
 RP PHOSPHORYLATION.  
 RC TISSUE=LIVER.  
 RX MEDLINE: 92331788.  
 RA Gillespie J.G., Hardie D.G.;  
 RT "Phosphorylation and inactivation of HMG-CoA reductase at the AMP-  
 RT activated protein kinase site in response to fructose treatment of  
 RT isolated rat hepatocytes.";  
 RL FEBS Lett. 306:59-62(1992).  
 CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE  
 CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING  
 CC ENZYME OF STEROL BIOSYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: (S)-MEVALONATE + COA + 2 NADP(+) = 3-HYDROXY-  
 CC 3-METHYLGUTARYL-COA + 2 NADPH.  
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC  
 CC RETICULUM.  
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M29249; AAA40608.1;  
 DR PFAM: PF00368; HMG-COA\_red. 1.  
 DR PRINTS: PR00071; HMGCOARDTASE.  
 DR PROSITE: PS00066; HMG\_COA\_REDUCTASE\_1; 1.  
 DR PROSITE: PS00318; HMG\_COA\_REDUCTASE\_2; 1.  
 DR PROSITE: PS01192; HMG\_COA\_REDUCTASE\_3; 1.  
 DR PROSITE: PS50065; HMG\_COA\_REDUCTASE\_4; 1.  
 KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;  
 KW Cholesterol biosynthesis; NADP; Phosphorylation.  
 FT DOMAIN 1 339  
 FT DOMAIN 340 449  
 FT DOMAIN 450 887  
 FT TRANSMEM 10 39  
 FT TRANSMEM 57 78  
 FT TRANSMEM 90 114  
 FT TRANSMEM 124 149  
 FT TRANSMEM 160 187  
 FT TRANSMEM 192 220  
 FT TRANSMEM 313 339  
 FT ACT\_SITE 558 558  
 FT ACT\_SITE 765 766  
 FT ACT\_SITE 865 865  
 FT CARBOHYD 281 281  
 FT CARBOHYD 517 517  
 FT CARBOHYD 869 869  
 FT MOD\_RES 871 871  
 PHOSPHORYLATION.

FT CONFLICT 423 423 S -> T (IN REF. 3).  
 FT CONFLICT 428 428 L -> V (IN REF. 3).  
 FT CONFLICT 443 443 P -> N (IN REF. 3).  
 SQ SEQUENCE 887 AA; 96592 MW; D941273D06573D22 CRC64;  
 Query Match 2.0%; Score 183; DB 1; Length 887;  
 Best Local Similarity 18.7%; Pred. No. 4.33e-12;  
 Matches 31; Conservative 68; Mismatches 62; Indels 5; Gaps 5;  
 Db 61 DIILITRCALIIYIYFQFNLRLQGLSGKYLGLAGLFTIFSSPFVS-TYVHF-LDKEL 118  
 Qy 305 ELVPTVAFMLVFAVYFVSVKRIDVFRSRFLALCSYTTTGSLMSGLCFEFTLSL 364  
 Db 119 TGLNEALPFLLLIDLSASALAKFALNSQDEYRENIARGMALIGPTFLDALVECLV 178  
 Qy 365 QS-KDIFVIVLVGLNLSLVTSSVSMDEFPVKIHAVALSSEGHISKLTLETLI 423  
 Db 179 IGVGTMSGVROLEIMCGCGSVLANFVETFFPACSLVLESR 224  
 Qy 424 LTIGLAFVPIVIOEFCIFALVGLSLDFMLQMLLFST-I-LAMNIRK 467  
 RESULT 13  
 ID HMDH\_RABBIT STANDARD: PRT: 888 AA.  
 AC Q29512;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA  
 DE REDUCTASE).  
 GN HMGCR.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEW ZEALAND WHITE; TISSUE=LIVER;  
 RA Yamada M., Yoshimatsu M., Kinowaki M., Kai M., Setoguchi T.;  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: THIS TRANSMEMBRANE GLYCOPROTEIN IS INVOLVED IN THE  
 CC CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT IS THE RATE-LIMITING  
 CC ENZYME OF STEROL BIOSYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-  
 CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.  
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC  
 CC RETICULUM AND PEROXISOMES.  
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.  
 DR PFAM: PF00368; HMG-COA\_red. 1.  
 DR PRINTS: PR00071; HMGCOARDTASE.  
 DR PROSITE: PS00066; HMG\_COA\_REDUCTASE\_1; 1.  
 DR PROSITE: PS00318; HMG\_COA\_REDUCTASE\_2; 1.  
 DR PROSITE: PS01192; HMG\_COA\_REDUCTASE\_3; 1.  
 DR PROSITE: PS50065; HMG\_COA\_REDUCTASE\_4; 1.  
 KW Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;  
 KW Cholesterol biosynthesis; NADP; Peroxisome.  
 FT DOMAIN 1 339  
 FT DOMAIN 340 449  
 FT DOMAIN 450 888  
 FT TRANSMEM 10 39  
 FT TRANSMEM 57 78  
 FT TRANSMEM 90 114  
 FT TRANSMEM 124 149  
 FT TRANSMEM 160 187  
 FT TRANSMEM 192 220  
 FT TRANSMEM 313 339  
 FT ACT\_SITE 558 559  
 FT ACT\_SITE 767 767  
 FT ACT\_SITE 866 866  
 FT CARBOHYD 281 281  
 FT CARBOHYD 296 296  
 FT CARBOHYD 518 518  
 LINKER.  
 CATALYTIC.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 GENERAL BASE (BY SIMILARITY).  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.





**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 NWSEIN (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:34:50 2000; Maspar time 39.26 Seconds

Tabular output not generated. 897.114 Million cell updates/sec

Title: >US-09-332-522B-4  
 Description: (1-508) from US09332522B.pep  
 Perfect Score: 3806  
 Sequence: 1 MDPFVFIVIASLYGVLYFF.....LIFALALIKVAMLSFLRPL 508

Scoring table: PAM 150  
 Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: spiremb112  
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
 13:sp-veterebrate 14:sp-virus

Statistics: Mean 50.285; Variance 102.588; scale 0.490

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	222	5.8	368	10	065444	1.15e-17
2	221	5.8	383	1	09YFP0	1.68e-17
3	212	5.6	377	1	057994	1.68e-17
4	193	5.1	389	1	027421	5.25e-16
5	175	4.6	385	1	P95972	6.50e-13
6	146	3.8	501	1	028947	4.51e-10
7	134	3.5	165	2	092868	9.98e-06
8	135	3.5	2846	5	09Y0A2	4.91e-04
9	135	3.5	2848	5	061193	3.57e-04
10	135	3.5	2857	5	09Y0A0	3.57e-04
11	135	3.5	2892	5	09Y0A1	3.57e-04
12	130	3.4	372	4	09Y5Y3	1.73e-03
13	130	3.4	376	4	043898	1.73e-03
14	125	3.3	4199	2	P74440	8.12e-03
15	122	3.2	325	13	042169	2.02e-02
16	120	3.2	344	8	078860	3.68e-02
17	123	3.2	715	5	09XUV0	1.49e-02
18	116	3.0	161	1	051981	1.20e-01
19	113	3.0	207	1	028302	2.85e-01
20	115	3.0	447	10	080885	1.60e-01

21	111	2.9	140	3	094594	VERY HYPOTHETICAL PROT	5.04e-01
22	111	2.9	199	10	P93296	ORE199.	5.04e-01
23	112	2.9	244	5	018352	C32C4.2 PROTEIN.	3.80e-01
24	112	2.9	326	13	042168	OLFACTORY RECEPTOR PRO	3.80e-01
25	112	2.9	487	2	033222	HYPOTHETICAL 52.6 KD P	3.80e-01
26	109	2.9	589	5	09X85	Y102ASC.28 PROTEIN.	8.86e-01
27	112	2.9	844	13	093552	POTATIVE ODORANT RECEPTOR	3.80e-01
28	110	2.9	5138	10	092W94	F5A8.4 PROTEIN.	6.69e-01
29	105	2.8	222	13	09W6V8	OLFACTORY RECEPTOR (FR	2.67e+00
30	107	2.8	232	1	09YCB9	232AA LONG HYPOTHETICAL	1.54e+00
31	108	2.8	266	11	061198	LY-49C.	1.17e+00
32	106	2.8	268	2	067364	HYPOTHETICAL 30.4 KD P	2.03e+00
33	106	2.8	283	2	092905	YNE.	2.03e+00
34	106	2.8	285	1	058347	285AA LONG HYPOTHETICAL	2.03e+00
35	108	2.8	326	13	042609	OLFACTORY RECEPTOR PRO	1.17e+00
36	106	2.8	326	13	042166	OLFACTORY RECEPTOR PRO	2.03e+00
37	105	2.8	326	13	042167	OLFACTORY RECEPTOR PRO	2.67e+00
38	106	2.8	344	8	079641	NADH DEHYDROGENASE SUB	2.03e+00
39	106	2.8	348	8	020949	NADH DEHYDROGENASE SUB	2.03e+00
40	105	2.8	372	13	P79945	HIGH-AFFINITY LYSOPHOS	2.67e+00
41	108	2.8	373	1	058089	373AA LONG HYPOTHETICAL	1.17e+00
42	106	2.8	470	4	043200	THYROID STIMULATING HO	2.03e+00
43	107	2.8	770	2	P73845	HYPOTHETICAL 87.8 KD P	1.54e+00
44	105	2.8	953	5	077046	AMINOPEPTIDASE N (EC 3	2.67e+00
45	106	2.8	1157	5	P90747	PROBABLE CALCIUM-TRANS	2.03e+00

## ALIGNMENTS

RESULT 1  
 ID 065444; PREDIMINARY; PRT; 368 AA.

AC 065444;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, last annotation update)  
 DE HYPOTHETICAL 40.3 KD PROTEIN.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA BEVAN M., TERRIN N., ARDILES W., BUYSHAERT C., DASSEVILLE R.,  
 RA DE CLERCK R., DE KEYSER A., MEYI P., ROUZE P., VAN DEN DAEL H.,  
 RA VILLAROEI R., GIELEN J., VAN MONTAGU M., BANCROFT I., MEWES H.W.,  
 RA MAYER K., SCHUELLER C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA ED ARABIDOPSIS SEQUENCING PROJECT;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL022224; CAAL18255.1; -;  
 KW PRINTS: PRO1000; SREBPS2PTASE.  
 DR HYPOTHETICAL protein.  
 SQ SEQUENCE 368 AA; 40315 MW; 036C342C CRC32;

Query Match 5.8%; Score 222; DB 10; Length 368;  
 Best local similarity 35.3%; Pred. No. 1.15e-17;  
 Matches 48; Conservative 34; Mismatches 46; Indels 8; Gaps 8;

Db	53	FS-CVLKVMVSIIGGFVAVSLIIVTVLLLOFHNPFSNRLTAV-FGPS-PSTRVSL	109
QY	78	FNWVL-VTFSL-LPDLI-LTIATIFSSGSDSSVSVGVQVLEILLPQVNP	134
Db	110	GIAVLVSTVIVSVHGLHAAASGIOMETAV-FIAIFGCVAFDNDVLS	168
QY	135	EIGYITIVLCLVYHMGHAAVMDVYTGIGINFICL-PLATLSHDLNLRW	193
Db	169	FNALRYCAGIWNVAV 184	
QY	194	FRKLRLVLCAGIWNFV 209	

ID	Accession	Species	Score	DB	Length	Indels	Gaps
RESULT	2	PRELIMINARY:	PRT:	383	AA.		
AC	09YEP0						
DT	01-NOV-1999	(TReMBLrel. 12, Created)					
DT	01-NOV-1999	(TReMBLrel. 12, Last sequence update)					
DT	01-NOV-1999	(TReMBLrel. 12, Last annotation update)					
DE	383AA LONG HYPOTHETICAL S2P METALLOPROTEASE.						
GN	AP0209.						
OS	Aeropyrum pernix.						
OC	Archaea; Crenarchaeota; Aeropyrum.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN-K1;						
RX	MEDLINE: 99310339.						
RA	KANARABAVASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,						
RA	JIN N. K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,						
RA	HOSOTAYAMA A., FUKUI S., NAGAI Y., NISHITAMA K., NAKAZAWA H.,						
RA	YAMAZAKI M., MASUDA S., FUNAHASHI T., TANAKA T., KUOCHI Y.,						
RA	TAAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,						
RA	NUMURA N., SAKO Y., KIKUCHI H.;						
RT	"Complete genome sequence of an aerobic hyper-thermophilic						
RL	crnarchaeon, Aeropyrum pernix K1."						
DR	DNA Res. 6:83-101(1999)						
DR	EMBL: AP000058; BAA9121.1;						
DR	Protease; Metalloprotease.						
SO	SEQUENCE 383 AA; 39729 MW; 4F8D3FEC CRC32;						
Query Match	5.8%; Score 221; DB 1; Length 383;						
Best Local Similarity	35.0%; Pred. No. 1,68e-17;						
Matches	36; Conservative 27; Mismatches 39; Indels 1; Gaps 1;						
Db	109 LIPGVTIEMEDLVYVAVALGAVYVAHELGHVAVVAVAGSIRKNGAIALIFLPAAYELD 168						
Oy	125 LLPGVNLPLEEIGYITTLVLCVLVHEHGMALAAVEDPVGTGIGIKFICPLPAYELSL 184						
Db	169 EDOLMKARLVSRKLVFSGVYANTILIALTL-LTAMAPPAEP 210						
Oy	185 HDHLSLRFMRKRLRYLCAGIWNEVFAGVCYLLISTVGIITMSP 227						
RESULT	3	PRELIMINARY:	PRT:	377	AA.		
ID	057994						
AC	057994;						
DT	01-AUG-1998 (TReMBLrel. 07, Created)						
DT	01-AUG-1998 (TReMBLrel. 07, Last sequence update)						
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)						
DE	377AA LONG HYPOTHETICAL PROTEIN.						
GN	PH0256.						
OS	Pyrococcus horikoshii.						
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN-OT3;						
RX	MEDLINE: 98344137.						
RA	KANARABAVASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,						
RA	YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAYAMA A., NAGAI Y.,						
RA	SAKAI M., OGURA K., OTOKA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,						
RA	FUNAHASHI T., TANAKA T., KUOCHI Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,						
RA	AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUDA N., SHIZUYA H.,						
RA	KIKUCHI H.;						
RT	"Complete sequence and gene organization of the genome of a hyper-						
RL	thermophilic archaeobacterium, Pyrococcus horikoshii OT3."						
DR	DNA Res. 5:55-76(1998)						
DR	EMBL: AP000001; BAA29328.1;						
DR	PFAM: PF00595; PDZ: 1.						
DR	PRINTS: PR01000; SREBP22PTASE.						
SO	SEQUENCE 377 AA; 41845 MW; 3FC8D3B2 CRC32;						
Query Match	5.6%; Score 212; DB 1; Length 377;						
Best Local Similarity	32.7%; Pred. No. 5.25e-16;						
Matches	37; Conservative 35; Mismatches 37; Indels 4; Gaps 4;						

Query	Subject	Score	E-value	Identical	Mismatches	Gaps
97	PFVPS-VOLVITPVTPLPV-YGL-IALAL-VLVEHSGFARSDIPKSGVGLLFI	152				
115	PVGVFVQLEILLPVPVNLPLEIEIGYITTVLCLVHEMGALAAVMDVYVTFGIGIFIF	174				
153	IIPAFVEPDEDQLKAPLRSLRFVFGAGSFANVALLISLLVNGIALAFEP	205				
175	CLPLATYELSHDHLNSLRWFRKLRLVLCAGIMHNVFVAGVGYLLISTVGTIMSP	227				
RESULT	4	PRELIMINARY	PRT	389	AA.	
ID	027421					
AC	027421					
DT	01-JAN-1998 (TREMBlrel. 05, Created)					
DT	01-JAN-1998 (TREMBlrel. 05, Last sequence update)					
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)					
DE	CONSERVED PROTEIN.					
GN	MTI1368.					
OS	Methanobacterium thermoautotrophicum.					
OC	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;					
CC	Methanobacterium.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=DELTA H:					
RX	MEDLINE: 98037514.					
RA	SMITH D.R., DOUGETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,					
RA	AUDREGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,					
RA	HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., OIU D.,					
RA	SPADAFORA R., VITCARE R., WANG Y., WIERZBOMSKI J., GIBSON R.,					
RA	JIYANI N., CARDOSO A., BUSH D., SAFER H., PATWEL D., PRABHAKAR S.,					
RA	MCDOUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,					
RA	DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;					
RT	"Complete genome sequence of Methanobacterium thermoautotrophicum					
RL	delta: functional analysis and comparative genomics."					
DR	J. Bacteriol. 179:7135-7155(1997).					
DR	EMBL: AE000899; AAB85845.1; -.					
DR	PFAM: PF00595; PDZ. 1.					
DR	PRINTS: PRO1000; SREBPS2PTASE.					
SQ	SEQUENCE 389 AA; 42956 MW; DC436662 CRC32;					
Query Match	5.1%;	Score 193;	DB 1;	Length 389;		
Best Local Similarity	29.5%;	Pred. No. 6.50e-13;				
Matches	28;	Conservative	34;	Mismatches 32;	Indels 1;	Gaps 1.
Db	113 LGYIGIGLAVTVYHEFAFGILARLEGVRIKISIGLLIIPGAFVPEDEDIKIRPIS	172				
Oy	136 IGYITTVLCLVYVHEMGALAAVMDVYVTFGIGIFIFCLPLATYELSHDHLNSLWFR	195				
Db	173 KMRIYAAGSVANLLIAGICFALFGISSFPAMPAF	207				
Oy	196 KLRVLCAGIMHNVFVAGVGY-LLISTVGTIMSP	229				
RESULT	5	PRELIMINARY	PRT	386	AA.	
ID	P95972					
AC	P95972					
DT	01-MAY-1997 (TREMBlrel. 03, Created)					
DT	01-MAY-1997 (TREMBlrel. 03, Last sequence update)					
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)					
DE	ORF C04034.					
OS	Sulfolobus solfataricus.					
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=P2.					
RX	MEDLINE: 97055432.					
RA	SENSEN C.W., KLENK H.P., SINGH R.K., ALLARD G., CHAN C.C.Y., LIU Q.Y.,					
RA	PENNY S.L., YOUNG F., SCHENK M.E., GAASSTERLAND T., DOOLITTLE W.F.,					
RA	RAGAN M.A., CHARLEBOIS R.L.;					
RT	"Organizational characteristics and information content of an archaeal					
RL	genome: 156 kb of sequence from Sulfolobus solfataricus P2."					
DR	MOL. Microbiol. 22:175-191(1996).					
DR	EMBL: Y08257; CAA69363.1; -.					





GN F33D4.2A.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;  
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BOERFELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GADNER A., GREEN P., HAWKINS J., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOMKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STILSON J.,  
RA THERRI-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA JOHNSON D., STELLYES L.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [13]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF036702; AAB88374.1; -  
DR PFAM: PF01365; RYDR-ITPR: 1.  
DR PRINTS: PRO0779; INSP3RECEPTR.  
SQ SEQUENCE 2848 AA; 325857 MW; DC91ABD4 CRC32;  
  
Query Match 3.5%; Score 135; DB 5; Length 2848;  
Best Local Similarity 26.3%; Pred. No. 3.57e-04;  
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;  
  
Db 2466 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVHVAFAVNSKGLD-DRPIAEIILAC 2520  
Oy 208 FVFAGVCYLLISTVGTITMSPLYAYNOHVYVELTRKSPLEGRGLQVDMQITQVNGC 264  
  
RESULT 10  
ID O9Y0A0 PRELIMINARY; PRT: 2857 AA.  
AC O9Y0A0.  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE INOSITOL 1,4,5-TRISPHOSPHATE RECEPTOR.  
GN ITR-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;  
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BAYLIS H.A., FURICHI T., YOSHIKAWA F., MIKOSHIBA K., SATTELLE D.B.;  
RA "Inositol 1,4,5-trisphosphate receptors are strongly expressed in the  
RA nervous system, pharynx, intestine and excretory cell of  
RA Caenorhabditis elegans and are encoded by a single gene (itr-1).";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ243182; CAB45863.1; -  
KW Receptor.  
SQ SEQUENCE 2857 AA; 327195 MW; CA316466 CRC32;  
  
Query Match 3.5%; Score 135; DB 5; Length 2857;  
Best Local Similarity 26.3%; Pred. No. 3.57e-04;  
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;  
  
Db 2475 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVHVAFAVNSKGLD-DRPIAEIILAC 2529

Oy 208 FVFAGVCYLLISTVGTITMSPLYAYNOHVYVELTRKSPLEGRGLQVDMQITQVNGC 264  
  
RESULT 11  
ID O9Y0A1 PRELIMINARY; PRT: 2892 AA.  
AC O9Y0A1.  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE INOSITOL 1,4,5-TRISPHOSPHATE RECEPTOR.  
GN ITR-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;  
OC Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BAYLIS H.A., FURICHI T., YOSHIKAWA F., MIKOSHIBA K., SATTELLE D.B.;  
RA "Inositol 1,4,5-trisphosphate receptors are strongly expressed in the  
RA nervous system, pharynx, intestine and excretory cell of  
RA Caenorhabditis elegans and are encoded by a single gene (itr-1).";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ243180; CAB45861.1; -  
KW Receptor.  
SQ SEQUENCE 2892 AA; 330832 MW; 148D2E24 CRC32;  
  
Query Match 3.5%; Score 135; DB 5; Length 2892;  
Best Local Similarity 26.3%; Pred. No. 3.57e-04;  
Matches 15; Conservative 26; Mismatches 14; Indels 2; Gaps 2;  
  
Db 2510 LILASLCELLSSIGVTLT-LYIFGILQLVNKIVHVAFAVNSKGLD-DRPIAEIILAC 2564  
Oy 208 FVFAGVCYLLISTVGTITMSPLYAYNOHVYVELTRKSPLEGRGLQVDMQITQVNGC 264  
  
RESULT 12  
ID O9Y5Y3 PRELIMINARY; PRT: 372 AA.  
AC O9Y5Y3.  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE ORPHAN G-PROTEIN-COUPLED RECEPTOR GPR45.  
GN GPR45.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE: 99156852.  
RA MARCHESE A., SAMZDARGO M., NGUYEN T., CHENG R., HENG H.H., NOWAK T.,  
RA IM D.S., LYNCH K.R., GEORGE S.R., O'DOWD B.F.;  
RA "Discovery of three novel orphan G-protein-coupled receptors.";  
RL Genomics 56:12-21(1999).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC O'DOWD B.F.;  
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF118266; AAD21056.1; -  
KW Receptor.  
SQ SEQUENCE 372 AA; 42000 MW; F8FA6D8E CRC32;  
  
Query Match 3.4%; Score 130; DB 4; Length 372;  
Best Local Similarity 27.4%; Pred. No. 1.73e-03;  
Matches 32; Conservative 33; Mismatches 44; Indels 8; Gaps 8;  
  
Db 19 SNADSGSTQLPAPRLSLAIVLMTVGF-LGNTVVCITVGR-PAMRSAT-NLLIAT 75  
Oy 108 SSSSVSPGVPOLEILPVPNLPLEIGIYITVLCVLVHMGHALAVMDVPTG 167  
  
Db 76 LAFSDIM-LSLCCAPFTAVTLITVRMHGDFHCLRSATLYWF-FVLEBVALILLISV 130  
Oy 168 FGKIFLPLAYTELSDHILNSLRW-F-RRLVLCAGI-WHNEFVAGVCYLLISTV 221

[illegible]

Query Match	3.3%	Score 125;	DB 2;	Length 4199;
Best Local	Similarity 28.6%;	Pred. No. 8.12e-03;		
Matches	34;	Conservative	30;	Mismatches 46;
			Indels	9;
			Gaps	8;

[illegible]

RESULT	15	PRELIMINARY:	PRT:	325 AA.
TD	042169			
AC	042169			
DT	01-JAN-1998	(TIREMBLrel, 05, Created)		
DT	01-JAN-1998	(TIREMBLrel, 05, Last sequence update)		
DT	01-NOV-1998	(TIREMBLrel, 08, Last annotation update)		
DE	OLFACTORY RECEPTOR PROTEIN 2.5			
OS	Brachydanio rerio (Zebrafish) (Zebra danio).			
OC	Euphyrotia, Metaceta; Chordata; Craniata; Vertebrata; Actinopterygii;			
OC	Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;			
OC	Cyprinoidae; Cyprinidae; Rasbora; Danio.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AB.			
RX	MEDLINE: 97436752.			
RA	BARTH A.L., DUGAS J.C., NGAI J.;			
RT	"Noncoordinant expression of odorant receptor genes tightly linked in			
RT	the zebrafish genome.";			
RL	Neuron 19:359-369(1997).			
DR	EMBL: AF012759: AAC60267.1; -.			
PF	PEAM; PF00001; 7tm_1; 1.			
QO	SEQUENCE 325 AA; 36842 MW; E6AAB7A CRC32;			

Search completed: Sun Sep 3 12:36:07 2000  
Job time : 77 secs.

Search completed: Sun Sep 3 12:36:07 2000  
Job time : 77 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
NWSEED (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_Lp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sun Sep 3 12:34:03 2000; Maspar time 17.06 Seconds  
922.625 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-332-522B-4  
Description: (1-508) from US09332522B.pep  
Perfect Score: 3806  
Sequence: 1 MDPFVEFVLASLYGLVFF.....LTFALALXVAMLSFRLPL 508

Scoring table: PAM 150  
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 51.577; Variance 96.756; scale 0.533

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	800	21.0	510	1	MS2P_CRIGR STEROL-REGULATORY ELEM	1.32e-138
2	723	19.0	519	1	MS2P_HUMAN STEROL-REGULATORY ELEM	6.10e-122
3	117	3.1	396	1	YDEA_ECOLI HYPOTHEICAL 42.5 KDA	2.11e-02
4	113	3.0	422	1	YF21_METUA PUTATIVE NA(+)/H(+) EX	7.24e-02
5	115	3.0	450	1	Y325_HAEIN HYPOTHEICAL PROTEIN H	3.92e-02
6	111	2.9	320	1	AA3R_RAT ADENOSINE A3 RECEPTOR	1.33e-01
7	109	2.9	641	1	NUSM_ALIMA NADH-UBIQUINONE OXIDOR	2.41e-01
8	111	2.9	1785	1	YCFL_ARATH HYPOTHEICAL 213.7 KDA	1.33e-01
9	108	2.8	138	1	GSPL_XANCP GENERAL SECRETION PATH	3.22e-01
10	107	2.8	224	1	ATP6_DROSI ATP SYNTHASE A CHAIN (	4.35e-01
11	108	2.8	266	1	KLR3_MOUSE KILLER CELL LECTIN-LIK	3.25e-01
12	107	2.8	360	1	CB2R_HUMAN CANNABINOID RECEPTOR 2	4.35e-01
13	107	2.8	459	1	VIPR_RAT VASOACTIVE INTESTINAL	4.35e-01
14	105	2.8	492	1	GTRL_MOUSE GLUCOSE TRANSPORTER TY	7.78e-01
15	105	2.8	492	1	GTRL_BOVIN GLUCOSE TRANSPORTER TY	7.78e-01
16	108	2.8	764	1	TSHR_RAT THYROTROPIN RECEPTOR P	5.83e-01
17	105	2.8	764	1	TSHR_HUMAN THYROTROPIN RECEPTOR P	5.83e-01
18	106	2.8	890	1	WES1_MOUSE WOLFGRAMIN	7.78e-01
19	106	2.8	1157	1	YES6_CAEEL PROBABLE CATION-TRANSP	5.83e-01
20	102	2.7	177	1	RM05_ACACA MITOCHONDRIAL 60S RIBO	1.83e+00
21	103	2.7	186	1	TRS2_ECOLI TRAS PROTEIN	1.38e+00
22	102	2.7	224	1	ATP6_DROME ATP SYNTHASE A CHAIN (	1.83e+00
23	101	2.7	226	1	ATP6_ANOOU ATP SYNTHASE A CHAIN (	2.42e+00

24	104	2.7	266	1	KLR8_MOUSE KILLER CELL LECTIN-LIK	1.04e+00
25	102	2.7	279	1	DM3C_HAEIN ANABROBIC DIMETHYL SUL	1.83e+00
26	102	2.7	296	1	Y639_SYNY3 HYPOTHEICAL 33.2 KDA	1.83e+00
27	103	2.7	298	1	MTRF_METKA TETRAHYDROMETHANOPTERIN	1.38e+00
28	101	2.7	313	1	OLF9_CANFA OLFACTORY RECEPTOR-LIK	2.42e+00
29	101	2.7	314	1	OLF9_RAT OLFACTORY RECEPTOR-LIK	2.42e+00
30	103	2.7	339	1	Y83C_METUA HYPOTHEICAL PROTEIN M	1.38e+00
31	104	2.7	351	1	CAV2_CAEEL CAVOOLIN-2	1.04e+00
32	102	2.7	352	1	NU2M_STRPU NADH-UBIQUINONE OXIDOR	1.83e+00
33	101	2.7	390	1	GTRL_SHEEP GLUCOSE TRANSPORTER TY	2.42e+00
34	102	2.7	450	1	VGLM_HAVER GLYCOPROTEIN M	1.83e+00
35	101	2.7	451	1	GTRL_PIG GLUCOSE TRANSPORTER TY	2.42e+00
36	102	2.7	492	1	GTRL_RABIT GLUCOSE TRANSPORTER TY	1.83e+00
37	101	2.7	492	1	GTRL_RAT GLUCOSE TRANSPORTER TY	2.42e+00
38	101	2.7	523	1	STC_RICCO SUGAR CARRIER PROTEIN	1.83e+00
39	102	2.7	541	1	LNT_RHIME APOLIPOPROTEIN N-ACYLT	1.83e+00
40	103	2.7	600	1	NUSM_METSE NADH-UBIQUINONE OXIDOR	1.38e+00
41	101	2.7	1103	1	CYCF_BOVIN RETINAL GUANYLYL CYCLA	2.42e+00
42	101	2.7	1133	1	Egf_RAT PRO-EPIDERMAL GROWTH F	2.42e+00
43	100	2.6	407	1	MOSC_RHIME MEMBRANE PROTEIN MOSC	3.20e+00
44	100	2.6	490	1	GTRL_CHICK GLUCOSE TRANSPORTER TY	3.20e+00
45	100	2.6	494	1	EXOT_RHIME SUCCINOGLYCAN BIOSYNTH	3.20e+00

## ALIGNMENTS

RESULT 1 STANDARD; PRT; 510 AA.

AC 054862: 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE STEROL-REGULATORY ELEMENT-BINDING PROTEINS INTRAMEMBRANE PROTEASE  
DE (EC 3.4.24.-) (SITE-2 PROTEASE).  
GN S2P.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98324087.  
RA Rawson R.B., Zelenksi N.G., Nijhawan D., Ye J., Sakai J., Hasan M.T.,  
RA Chang T.Y., Brown M.S., Goldstein J.L.;  
RT "Complementation cloning of S2P, a gene encoding a putative  
RT metalloprotease required for intramembrane cleavage of SREBPs";  
RL Mol. Cell 1:47-57(1997).  
CC -!- FUNCTION: INTRAMEMBRANE PROTEOLYSIS OF STEROL-REGULATORY ELEMENT-  
CC BINDING PROTEINS (SREBPS) WITHIN THE FIRST TRANSMEMBRANE SEGMENT  
CC THEREBY RELEASING THE N-TERMINAL SEGMENT WITH A PORTION OF THE  
CC TRANSMEMBRANE SEGMENT ATTACHED. SITE-2 CLEAVAGE COMES AFTER SITE-1  
CC CLEAVAGE WHICH TAKE PLACES IN THE LUMENAL LOOP.  
CC -!- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50 (ZINC  
CC METALLOPROTEASE).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF019611; AAC53526.1; -  
DR PRINTS: PRO1000; SREBPS2PTASE.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; zinc; Transmembrane.  
FT METAL 162 162 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 163 163 BY SIMILARITY.  
FT METAL 166 166 ZINC (CATALYTIC) (BY SIMILARITY).  
FT TRANSMEM 4 24 POTENTIAL.

Query Match 21.0%; Score 800; DB 1; length 510;  
 Best Local Similarity 30.0%; Pred. No. 1,32e-138;  
 Matches 155; Conservative 146; Mismatches 190; Indels 25; Gaps 19;

DB	Accession	Score	Length	Indels	Gaps
FT	TRANSMEM 75	95	POTENTIAL.		
FT	TRANSMEM 177	197	POTENTIAL.		
FT	TRANSMEM 220	240	POTENTIAL.		
FT	TRANSMEM 429	449	POTENTIAL.		
FT	TRANSMEM 494	504	POTENTIAL.		
FT	DOMAIN 109	127	POLY-SER.		
FT	DOMAIN 371	375	POLY-SER.		
SO	SEQUENCE 510 AA	56498 MW	1286B2257C0DA7A4 CRC64;		

Query Match 21.0%; Score 800; DB 1; length 510;  
 Best Local Similarity 30.0%; Pred. No. 1,32e-138;  
 Matches 155; Conservative 146; Mismatches 190; Indels 25; Gaps 19;

1 MIPVSLVVVVGWAVYVYADLVKSSVYFKHSYEDMLEKNGISLSPHIRMOTAVENRA 60  
 1 MDPFVFYVILASLYGVLYFFDRFKSCMHY--PYDAFLNKGISLSPHIRMOTAVENRA 58

61 FYSWGRKRAKMLYQWPFNGVYVAMFSSFFLLGKTLMOQLAOMADSPSSSSSSSS 120  
 59 LRMGSAGNSCT-RRYMIT-SFNV-GVLVTFSLPIGLI--L-LIA--TIFSSGEDSSS 110

121 SSSSSSSINEOVLYVVGINLPVNOIYFPAVILSGVHEIGHGIAIREOVFNGF 180  
 111 SVSSSPVGPV-Q-LEILLPGVNLPLEIGIYITTLVLCVHEHGHAAVMDVPTG 168

181 GI-FLFIIRPAFVDFTTHLOLISPVQOLRIFCAGIMHNFVALLGIALVLLPYLLP 239  
 169 GIKFITCL-PLATVETSHDNLNLRWFKRLVLCAGIMHNFVAGCYLLISVGTIMSP 227

240 FY-YTGVGLITEVAEDSPAIGRGLVGDVLTHTDCCPVTVNOVNECDLTATPEQIG 298  
 228 LVAYNOH-VVTELTRKSPLRGEGLOVDNQITQVNGCCPVNSESSESVTCION-SLKLRPG 285

299 YCISASTLOOLS-FPRRAKRLDGTSCCNHSLTDVCFEYRNKRLHTCIPARKAVE 357  
 286 YCVSAFVOLNDESSAISHSIDGOLCCDELNPVNSCFEYVADANGDVVELPQHVCLN 345

358 AVO-VCRNKDKTS-SSSFCIVPSLETETRLIKVHNP-QIDML---YVGHPLHLAY 410  
 346 VRTLEVESEHSSGVNCFCLPLIRNTAITMTRKROFRBKLPVLYVGHVMDVTR 405

411 TVSTSFIRFNFSLDLPVIVETFKYLLISGALAIYNAVPCFALDGMILNSFLDAT 470  
 406 IVEVSAFVPRYSLKAMPAMALLKYNVVFISGILALINAIAPFCGDAHITSTVISHP 465

471 LRSVIGDNVYKDLIGFILLGGSVLLAANTLELWM 506  
 466 LVGRVDOHAKRDIISLITSVGSLFALALIKYAML 501

RESULT 2  
 ID MS2P\_HUMAN STANDARD: PRT; 519 AA.  
 AC 043462;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE STEROL-REGULATORY ELEMENT-BINDING PROTEINS INTRAMEMBRANE PROTEASE  
 DE (EC 3.4.24.-) (SITE-2 PROTEASE).  
 GN 52P.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FIBROBLAST;  
 RX MEDLINE: 98324087.  
 RA Rawson R.B., Zelenski N.G., Nijhawan D., Ye J., Sakai J., Hasan M.T.,  
 RA Chang T.Y., Brown M.S., Goldstein J.L.;  
 RT "Complementation cloning of S2P, a gene encoding a putative  
 RT metalloprotease required for intramembrane cleavage of SREBPs.";  
 RL Mol. Cell 1:47-57(1997).  
 CC -1- FUNCTION: INTRAMEMBRANE PROTEOLYSIS OF STEROL-REGULATORY ELEMENT-  
 CC BINDING PROTEINS (SREBPs) WITHIN THE FIRST TRANSMEMBRANE SEGMENT

CC THEREBY RELEASING THE N-TERMINAL SEGMENT WITH A PORTION OF THE  
 CC TRANSMEMBRANE SEGMENT ATTACHED. SITE-2 CLEAVAGE COMES AFTER SITE-1  
 CC CLEAVAGE WHICH TAKE PLACES IN THE LUMENAL LOOP.  
 CC -1- COPACITOR: BINDS ONE ZINC ION (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,  
 CC LIVER, MUSCLE, KIDNEY AND PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50 (ZINC  
 CC METALLOPROTEASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: AF019612; AAC51937.1.-.  
 CC PRINTS: PRO1000; SREBPS2PTASE.  
 CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Zinc; Transmembrane.  
 CC METAL 171 171  
 CC ACT\_SITE 172 172  
 CC METAL 175 175  
 CC TRANSMEM 4 24  
 CC TRANSMEM 75 95  
 CC TRANSMEM 186 206  
 CC TRANSMEM 229 249  
 CC TRANSMEM 448 468  
 CC TRANSMEM 493 513  
 CC DOMAIN 109 136  
 CC DOMAIN 380 384  
 CC MUTAGEN 171 171  
 CC MUTAGEN 172 172  
 CC MUTAGEN 172 172  
 CC MUTAGEN 172 172  
 CC MUTAGEN 175 175  
 CC SEQUENCE 519 AA; 57443 MW; 247D69E0FD747BD CRC64;

Query Match 19.0%; Score 723; DB 1; length 519;  
 Best Local Similarity 29.9%; Pred. No. 6.10e-122;  
 Matches 157; Conservative 146; Mismatches 188; Indels 34; Gaps 20;

1 MIPVSLVVVVGWAVYVYADLVKSSVYFKHSYEDMLEKNGISLSPHIRMOTAVENRA 60  
 1 MDPFVFYVILASLYGVLYFFDRFKSCMHY--PYDAFLNKGISLSPHIRMOTAVENRA 58

61 FYSWGRKRAKMLYQWPFNGVYVAMFSSFFLLGKTLMOQLAOMADSPSSSSSSSS 120  
 59 LRMGSAGNSCT-RRYMIT-SFNV-GVLVTFSLPIGLI--L-LIA--TIFSSGEDSS 108

121 SSS 180  
 109 SSSVSSPVGPV---V---Q-LEILLPGVNLPLEIGIYITTLVLCVHEHGHAAV 159

181 REQVRNFGFI-FLFIIRPAFVDFTTHLOLISPVQOLRIFCAGIMHNFVALLGIAL 239  
 160 MEDVPTVGTGKIFITCL-PLATVETSHDNLNLRWFKRLVLCAGIMHNFVAGCYLLI 218

240 VLLPVLLPPT-YTGVGLITEVAEDSPAIGRGLVGDVLTHTDCCPVTVNOVNECD 298  
 219 STVGITMSPLYAYNOH-VVTELTRKSPLRGEGLOVDNQITQVNGCCPVNSESSESVTCIO 277

299 TIATPEQIGCIASASTLOOLS-FPRRAKRLDGTSCCNHSLTDVCFEYRNKRLHT 357  
 278 N-SLKLRPGCVSADVQOLNDESSAISHSIDGOLCCDELNPVNSCFEYVADANGDV 336

358 CLPARKAVEATO-VCRNKDKTS-SSSFCIVPSLETETRLIKVHNP-QIDML---Y 410  
 337 ELPOHVCANRLILEVESEHSSGVNCFCLPLIRNTAITMTRKROFRBKLPVLY 396

411 VGHPLHLAYTVSTSFIRFNFSLDLPVIVETFKYLLISGALAIYNAVPCFALDGM 470



```

FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 378 398 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
SQ SEQUENCE 450 AA; 47439 MW; 4FD139E73FDD8061 CRC64;

Query Match 3.0%; Score 115; DB 1; Length 450;
Best Local Similarity 25.7%; Pred. No. 3,92e-02;
Matches 19; Conservative 24; Mismatches 28; Indels 3; Gaps 3;

Db 166 IGMILGLTATFIT 219
Oy 71 RRAVACVLTGEGTATYILPGVFGKGFISTILVKININAGATLGLQTNAVQSLAMLPV 205
Oy 71 RRAVATTSNNGVLTFTSLPIGL-ILLATITFSSEGDSSSSVSPGV-P-VQLEILLPG 128
Oy 129 VNLPLEEI-GYYIT 141

RESULT 6 STANDARD; PRT; 320 AA.
ID AA3R_RAT
AC P28647; Q63792;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE ADENOSINE A3 RECEPTOR (TGPR1).
GN ADORA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-BRAIN.
RX MEDLINE; 92366475.
RA Zhou Q.Y., Li C.Y., Olah M.E., Johnson R.A., Stiles G.L., Civelli O.;
RT "Molecular cloning and characterization of an adenosine receptor: the
RT A3 adenosine receptor.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:7432-7436(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-TESTIS;
RX MEDLINE; 91285122.
RA Meyerhof W., Mueller-Brechlin R., Richter D.;
RT "Molecular cloning of a novel putative G-protein coupled receptor
RT expressed during rat spermiogenesis.";
RT FEBS Lett. 284:155-160(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 96196578.
RA Sajjadi F.G., Boyle D.L., Domingo R.C., Firestein G.S.;
RT "cDNA cloning and characterization of A1i, an alternatively spliced
RT rat A3 adenosine receptor variant.";
RT FEBS Lett. 382:125-129(1996).
RC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
IS MEDIATED BY G PROTEINS WHICH INHIBITS ADENYL CYCLASE.
CC POSSIBLE ROLE IN REPRODUCTION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE
SPlicing.
CC -1- TISSUE SPECIFICITY: TESTIS, PARTICULARLY IN SPERMATOCTYES AND
SPERMATIDS BUT NOT IN SPERMATOGENIA. LOW LEVELS IN THE BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING SPERMIOGENESIS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC

```



```
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; X59249; CAA41937.1; -
DR EMBL; M94152; AAA0680.1; -
DR EMBL; X93119; CAA63702.1; -
DR PIR; S17177; S17177.
DR PIR; A46152; A46152.
DR HSSP; P29274; IMMH.
DR GCRDB; GCR_0177; -.
DR GCRDB; GCR_0373; -.
DR GCRDB; GCR_1362; -.
DR PFAM; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PRINTS; PR00424; ADENOSTINER.
DR PRINTS; PR00555; ADENOSTINEA3R.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Lipoprotein; Palmitate; Alternative splicing.
KW DOMAIN; 1
FT TRANSMDM 1 16
ET EXTRACELLULAR (POTENTIAL).
FT TRANSMDM 17 39
ET DOMAIN 40 50
FT TRANSMDM 51 74
ET DOMAIN 75 86
FT TRANSMDM 87 108
ET DOMAIN 109 128
FT TRANSMDM 129 150
FT TRANSMDM 151 179
FT TRANSMDM 180 200
FT TRANSMDM 201 233
FT TRANSMDM 234 257
FT TRANSMDM 258 263
FT TRANSMDM 264 286
FT DOMAIN 287 320
FT CARBOHYD 4
FT CARBOHYD 5
FT DISULFID 85
FT LIPID 305
FT VASAPLIC 119 119
FT CONFLICT 18 18
FT CONFLICT 68 68
FT CONFLICT 74 83
FT CONFLICT 132 132
FT CONFLICT 172 172
FT CONFLICT 288 288
FT CONFLICT 291 295
SO SEQUENCE 320 AA; 35629 MW; 59DBA40C94ACDAB CRC64;

Query Match 2.9%; Score 111; DB 1; Length 320;
Best Local Similarity 27.4%; Pred. No. 1,33e-01;
Matches 32; Conservative 34; Mismatches 39; Indels 12; Gaps 10;

Db 51 YEIVSLADLVAGVGLVPLAIYASLEVMHRYACLFMSCVLLVTFTHSIMSLIAIADR 110
Oy 138 YIIITVLVC-LIVHEKMGHALA-AVEDVDPVVGFGIKFFCLDELTATLSH-DHLN-St-R 192
Db 111 YLRVALTYRYRTVTFQRRIWILFLGCMWL-VSEFL-VGLTIPMFQMNKRVTL-ELSONSS 164
Oy 193 WFR-KLRRLCAGI---WHNFVFAGVCYILLISTVGIIMSPLYANQHVVYTELTRKSP 245

RESULT 7
ID NUSM_ALIMA STANDARD: PRT: 641 AA.
AC P50365;
DT 01-OCT-1996 (Rel. 34, Created)
DI 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NAOH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN ND5 OR NAD5.
OS Allomyces macrogynus.
OG Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Blastocladiiales; Blastocladiaceae;
```

```
CC Allomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
RA Paquin B., Roewer I., Wang Z., Lang B.F.;
RT "A robust fungal phylogeny using the mitochondrially encoded nad5
protein sequence.";
RL Can. J. Bot. 73:S180-S185(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 46923 / BURMA 3-35 (350C);
RX MEDLINE: 96226032.
RA Paquin B., Lang B.F.;
RT "The mitochondrial DNA of Allomyces macrogyrus: the complete genomic
sequence from an ancestral fungus.";
RL J. Mol. Biol. 255:688-701(1996).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U17010; AAB05846.1; -.
DR EMBL: U41288; AAC49228.1; -.
DR PDBM: PF00361; oxidored_q1.N: 1.
DR PDBM: PF00662; oxidored_q1.N: 1.
KW Oxidoreductase; NAD; ubiquinone; Mitochondrion.
SQ SEQUENCE 641 AA; 70674 MW; 9664c378b72e7be1 CRC64;
Query Match 2.9%; Score 109; DB 1; Length 641;
Best Local Similarity 29.4%; Pred. No. 2,41e-01;
Matches 15; Conservative 15; Mismatches 20; Indels 1; Gaps 1;
Db Db 141 GWEGIGVSSYLINFWFTRI-QANKASQAMLVNRVGDLGFVLGFGIFWL 190
OY 451 GFDEAHITSTVFHSFLRGVDQHAKRDIISLIITSVGLPALMLKYAWL 501
RESULT 8 STANDARD; PRT; 1786 AA.
AC YCF1_ARATH STANDARD; PRT; 1786 AA.
AC P56785;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 2.3.7 KDA PROTEIN ycf1.
GN YCF1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsids.
CN [1]
RN RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
RC STRAIN=cv. COLOMBIA;
RT "Complete structure of the chloroplast genome of Arabidopsis
thaliana.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC -I- FUNCTION: NOT YET KNOWN.
CC -I- SIMILARITY: TO A SIMILAR ORF IN OTHER PLANTS CHLOROPLAST.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
```



OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB.17;  
RX MEDLINE: 95355830.  
RA Stoneman E.R., Bennett M., An J., Chesnut K.A., Wakeford E.K.,  
RA Schreier J.B., Scigliano M.J., Kumar V., Mathew P.A.;  
RT "Cloning and characterization of 5E6(ly-49c), a receptor molecule  
RT expressed on a subset of murine natural killer cells.";  
RL J. Exp. Med. 182:305-313(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=LUNG;  
RX MEDLINE: 91332459.  
RA Wong S., Freeman J.D., Kelleher C., Mager D., Takei F.;  
RT "Ly-49 multigene family. New members of a superfamily of type II  
RT membrane proteins with lectin-like domains.";  
RL J. Immunol. 147:1417-1423(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6, AND 129/SVJ;  
RX MEDLINE: 95356819.  
RA Held W., Roland J., Raullet D.H.;  
RT "Allelic exclusion of Ly49-family genes encoding class I MHC-specific  
RT receptors on NK cells.";  
RL Nature 376:355-358(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/SN;  
RA Sundback J., Karre K., Sentman C.L.;  
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=LUNG;  
RA Wong S., Freeman J.D., Kelleher C., Mager D., Takei F.;  
RL Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C, NZB, C57BL/6, AND BALB/C X C57BL/6;  
RA Mathew P.A., Stoneman E., Bennett M., An J., Chesnut K.A.,  
RA Wakeford E.K., Schreier J.B., Scigliano M.J., Kumar V.;  
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.  
CC -!- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR CLASS I MHC.  
CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by, and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U09739; AAA86873.1; -;  
DR EMBL: U49865; AAA92851.1; -;  
DR EMBL: U49866; AAA92852.1; -;  
DR EMBL: U49867; AAA92953.1; -;  
DR EMBL: U49868; AAA92954.1; -;  
DR EMBL: U10305; AAA19053.1; -;  
DR EMBL: U56405; AAB19101.1; -;  
DR EMBL: U34891; AAA77065.1; -;  
DR EMBL: U34892; G1003009; -;  
DR HSSP: P04511; 1LIT.  
DR MGD: MGI:101905; KLRK3.  
DR PFAM: PF00059; lectin\_c\_1.  
DR PROSITE: PS00615; C-TYPE-LECTIN\_1; FALSE\_NEG.  
DR PROSITE: PS50041; C-TYPE-LECTIN\_2; 1  
DR T-cell; Glycoprotein; Antigen; Transmembrane; Cell adhesion;  
KW Signal-anchor; Lectin; Receptor; Multigene family.  
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 45 66 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 67 266  
FT DOMAIN 143 261  
FT DISULFID 154 257  
FT DISULFID 171 249  
FT CARBOHYD 79 79  
FT CARBOHYD 87 87  
FT CARBOHYD 104 104  
FT CARBOHYD 113 113  
FT CARBOHYD 160 160  
FT VARIANT 2 2  
FT VARIANT 22 22  
FT VARIANT 34 34  
FT VARIANT 42 42  
FT VARIANT 60 60  
FT VARIANT 65 66  
FT VARIANT 72 72  
FT VARIANT 85 85  
FT VARIANT 93 93  
FT VARIANT 115 115  
FT VARIANT 117 117  
FT VARIANT 127 127  
FT VARIANT 129 129  
FT VARIANT 131 131  
FT VARIANT 133 133  
FT VARIANT 146 146  
FT VARIANT 151 151  
FT VARIANT 174 174  
FT VARIANT 179 179  
FT VARIANT 189 189  
FT VARIANT 198 198  
FT VARIANT 219 219  
FT VARIANT 226 226  
FT VARIANT 232 232  
FT VARIANT 247 247  
FT VARIANT 251 251  
FT VARIANT 260 260  
FT SEQUENCE 266 AA: 31285 MM: 409F61E5DFA2299A CXC64:  
Query Match 2.8%; Score 108; DB 1; Length 266;  
Best Local Similarity 38.9%; Pred. No. 3, 25e-01;  
Matches 14; Conservative 11; Mismatches 9; Indels 2; Gaps 2;  
DB 39 CSAFWOLYKALGICFLLYLVANVLAKITOYNQH 74  
QY 201 CAGIWHNFVFA-GV-CYLLISTVGTMSPLAYAYNOH 234  
RESULF 12 STANDARD: PRT: 360 AA.  
AC CB2B\_HUMAN  
AC P34972;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)

```
DE CANNABINOID RECEPTOR 2 (CB2) (CB-2) (CX5).
GN CNR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93368659.
RA Munro S., Thomas K.L., Abu-Shaar M.;
RT "Molecular characterization of a peripheral receptor for
RL cannabinoids".
RI Nature 365:61-65(1993).
CC CC -I- FUNCTION: INVOLVED IN CANNABINOID-INDUCED CNS EFFECTS. COULD BE A
CC CC RECEPTOR FOR ANANDAMIDE.
CC CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC CC -I- TISSUE SPECIFICITY: NOT FOUND IN BRAIN. FOUND IN SPLEEN.
CC CC -I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage is for commercial
CC entities requires a license agreement (See http://www.isb.slb.ch/announce/
CC or send an email to license@isb.slb.ch).
-----
DR EMBL: X74328; CAA52376.1; -.
DR GCRDB: GCR_0645; -.
DR PFMAT: PF00001; 7tm.1; 1.
DR PRINTS: PR00362; CANNABINOIDR.
DR PRINTS: PR00523; CANNABINOIDR.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
KW G-Protein coupled receptor; Transmembrane; glycoprotein.
FT DOMAIN 1 33
FT TRANSMEM 34 59
FT DOMAIN 60 71
FT TRANSMEM 72 92
FT DOMAIN 93 104
FT TRANSMEM 105 129
FT DOMAIN 130 149
FT TRANSMEM 150 172
FT DOMAIN 173 188
FT TRANSMEM 189 214
FT DOMAIN 215 246
FT TRANSMEM 247 267
FT DOMAIN 268 279
FT TRANSMEM 280 301
FT DOMAIN 302 360
FT CARBOHD 11 11
SQ SEQUENCE 360 AA; 39680 MW; A7BCF68C16E7514B CR664;

Query Match 2.8%; Score 107; DB 1; Length 360;
Best Local Similarity 23.5%; Pred. No. 4,35e-01;
Matches 39; Conservative 49; Mismatches 64; Indels 14; Gaps 12;

Db 129 IDRLV--CLRVPPSKALLTR-GRALVTGIMWVLASLVSYPLMGWTCPCPRCSLPFL 185
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
20 FDRFKSCMHYP--YDAFLAKNTGLSLNPSLSLMHHSIAFNRTL-LR WSGAGNSCTRR-V 74
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 IPNDLYLSWLFIKFAELFSGIITTYGHVLMKAHQHAS-LSGHODROQPGMARRLDVRLA 244
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
75 ITS-FNVGVLLVFSLPIGLILLIILATIIFSSGDSDSSSVSPGVVQ-LEILLPENVLP 132
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 245 -KTLDVLAVLLICWFPPVVALAHSLATLTLSQVKAFPCSLCL 289
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 133 LEELISYITTLVLC-LVAHE-MGHALAAMEDEVPTVGFIKFIEFL 176
OY :||: ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
ID VIPR_RAT STANDARD; PRT; 459 AA.
AC P30083;
DT 01-APR-1993 (Rel. 25, Created)
ZT 01-APR-1993 (Rel. 25, Last sequence update)
```

DT 01-NOV-1997 (Rel. 35, last annotation update)  
 DE VASOACTIVE INTENSINAL POLYPEPTIDE RECEPTOR 1 PRECURSOR (VIP-R-1)  
 DE (PUTATIVE ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE II RECEPTOR)  
 DE (PACAP TYPE II RECEPTOR) (PACAP-R-2).  
 GN VIPR1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG.  
 RX MEDLINE; 92232309.  
 RA "Inuhara T., Shigemoto R., Mori K., Takahashi K., Nagata S.:  
 RT "Functional expression and tissue distribution of a novel receptor  
 RT for vasoactive intestinal polypeptide".  
 RL Neuron 8:811-819(1992).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR VIP. THE ACTIVITY OF THIS  
 CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLIL  
 CC CYCLASE.  
 CC -1- SOURCE/CELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: IN LIVER, LUNG, INTESTINES, THYMUS AND BRAIN  
 CC (MOSTLY IN THE CEREBRAL CORTEX AND HIPPOCAMPUS).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaborat  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on i  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commerc  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M66835; AAA42331.1; -.  
 DR PIR; JH0594; JH0594.  
 DR GCRDB; GCR\_0369; -.  
 DR PFAM; PF00002; 7tm\_2; 1.  
 DR PRINTS; PR00049; GPCRSECRETIN.  
 DR PRINTS; PR00491; VASOACTIVEIPR.  
 DR PRINTS; PR01154; VIPRRECEPTOR.  
 DR PROSITE; PS00649; G-PROTEIN\_RECP\_F2\_1; 1.  
 DR PROSITE; PS00650; G-PROTEIN\_RECP\_F2\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1  
 FT CHAIN 31 459  
 FT  
 FT DOMAIN 31 143  
 FT TRANSSEM 144 168  
 FT DOMAIN 169 175  
 FT TRANSSEM 176 195  
 FT DOMAIN 196 217  
 FT TRANSSEM 218 241  
 FT DOMAIN 242 255  
 FT TRANSSEM 256 277  
 FT DOMAIN 278 294  
 FT TRANSSEM 295 318  
 FT DOMAIN 319 343  
 FT TRANSSEM 344 363  
 FT DOMAIN 364 375  
 FT TRANSSEM 376 395  
 FT DOMAIN 396 459  
 FT CARBOHYD 58  
 FT CARBOHYD 69  
 FT CARBOHYD 100 100  
 FT CARBOHYD 292  
 FT SEQUENCE 459 AA; 52057 MW; 99E857DA86698D2 CRC64;  
 Query Match 2.8%; Score 107; DB 1; Length 459;  
 Best Local Similarity 22.4%; Pred.No. 4,35e-01;  
 Matches 30; Conservative 40; Mismatches 54; Indels 10; Gaps 9

```

Db      283 DFGCWDITINSSLMWIKAPILSLVNFVLF-ICIRILVOKRPPDICKNDSSPYSL 341
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :
Cc      58 TLRWMSGAGSCTRHWMTITSNGVLVTFSLPIGLI-LLIATIFSSGEODSSSVSPV 116
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :
Db      342 AKSTLLIPLLEGIH 355
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :
Cc      117 GVPVQLLEILPGVN 130
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :

RESULT 14
ID      GTRL_MOUSE      STANDARD:      PRT:      492 AA.
AC      P17809; Q61608;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      GLUCOSE TRANSPORTER TYPE 1, ERYTHROCYTE/BRAIN (Gtl).
GN      SLC2A1 OR GLUT1 OR GLUT-1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-BRAIN;
RX      MEDLINE: 89240694.
RA      Kaestner K.H., Christy R.J., McLeathlan J.C., Braiterman L.T.,
RT      Cornelius P., Pekala P.H., Lane M.D.;
RT      "Sequence, tissue distribution, and differential expression of mRNA
RT      for a putative insulin-responsive glucose transporter in mouse 3T3-L1
RT      adipocytes."
RL      Proc. Natl. Acad. Sci. U.S.A. 86:3150-3154(1989).
RN      [2]
RP      MEDLINE: 90274408.
RX      Reed B.C., Shade D., Alperovich F., Yang M.;
RT      "3T3-L1 adipocyte glucose transporter (HepG2 class): sequence and
RT      regulation of protein and mRNA expression by insulin,
RT      differentiation, and glucose starvation."
RL      Arch. Biochem. Biophys. 279:261-274(1990).
RN      [3]
RP      SEQUENCE OF 151-237 FROM N.A.
RC      STRAIN-C57BL/6; TISSUE-EMBRYO;
RX      MEDLINE: 93170163.
RA      Smith D.E., Gridley T.;
RT      "Differential screening of a PCR-generated mouse embryo cDNA library:
RT      glucose transporters are differentially expressed in early
RT      postimplantation mouse embryos."
RL      Development 116:555-561(1992).
RN      [4]
RP      SEQUENCE OF 357-463 FROM N.A.
RX      MEDLINE: 92111400.
RA      Hogan A., Heyner S., Charbon M.J., Copeland N.G., Gilbert D.J.,
RA      Jenkins N.A., Thorens B., Schultze G.A.;
RT      "Glucose transporter gene expression in early mouse embryos."
RL      Development 113:363-372(1991).
Cc      -!- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM MAY BE
Cc      RESPONSIBLE FOR CONSTITUTIVE OR BASAL GLUCOSE UPTAKE. HAS A VERY
Cc      BROAD SUBSTRATE SPECIFICITY; CAN TRANSPORT A WIDE RANGE OF ALDOSES
Cc      INCLUDING BOTH PENTOSE AND HEXOSES.
Cc      -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
Cc      -!- DEVELOPMENTAL STAGE: LEVELS DECLINE 3-FOLD BETWEEN DAYS 7.5 AND
Cc      12.5 OF GESTATION. AT 7.5 DPC, EXPRESSED MORE STRONGLY IN
Cc      EXTREMEZYOTIC TISSUES THAN IN THE EMBRYO PROPER. EXPRESSED IN
Cc      ANIONIC, CHORION, AND ECTOPLACENTAL CONE. IN THE YOLK SAC,
Cc      EXPRESSED MORE STRONGLY IN THE MESODERM LAYER THAN THE ECTODERM.
Cc      EXPRESSION FAIRLY WIDESPREAD IN THE EMBRYO AT 8.5 DPC, BUT BY 10.5
Cc      DPC, EXPRESSION IS DOWN-REGULATED AND OBSERVED IN THE EYE AND THE
Cc      SPINAL CORD.
Cc      -!- SIMILARITY: HIGH TO OTHER GLUCOSE TRANSPORTERS.
Cc      -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
Cc      -----
Cc      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.isb.ch/announce/
Cc      or send an email to license@isb.sib.ch).
Cc      -----
DR      EMBL: M22998; AAA37707.1; -
DR      EMBL: M23384; AAA37752.1; -
DR      EMBL: S77924; CAB33671.1; -
DR      EMBL: X69697; CAA49367.1; -
DR      PIR: A30310; A30310.
DR      PIR: S09705; S09705.
DR      HSP: P00844; 1A91.
DR      MGD: MGI:95755; SLC2A1.
DR      PFAM: PF00083; sugar tr: 1.
DR      PRINTS: PR00171; SUGRTNSPORT.
DR      PRINTS: PR01172; GLUCTRNSPORT.
DR      PRINTS: PR01190; GLUCTRNSPORT.
DR      PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
DR      PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW      Duplication; Transmembrane; Sugar transport; Transport; Glycoprotein;
KW      Multigene family.
FT      TRANSMEM 13 33 POTENTIAL.
FT      DOMAIN 34 66 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 67 87 POTENTIAL.
FT      TRANSMEM 96 116 POTENTIAL.
FT      TRANSMEM 127 147 POTENTIAL.
FT      TRANSMEM 156 176 POTENTIAL.
FT      TRANSMEM 186 206 POTENTIAL.
FT      DOMAIN 207 271 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 272 292 POTENTIAL.
FT      TRANSMEM 307 327 POTENTIAL.
FT      TRANSMEM 338 358 POTENTIAL.
FT      TRANSMEM 369 389 POTENTIAL.
FT      TRANSMEM 402 422 POTENTIAL.
FT      TRANSMEM 430 450 POTENTIAL.
FT      DOMAIN 451 492 CYTOPLASMIC (POTENTIAL).
FT      CARBOHYD 45 45 POTENTIAL.
FT      CONFLICT 52 52 I -> Y (IN REF. 2).
FT      CONFLICT 193 195 VFV -> Q (IN REF. 2).
FT      CONFLICT 358 358 L -> Q (IN CAB33671).
FT      CONFLICT 360 360 R -> Q (IN REF. 4).
FT      CONFLICT 403 403 A -> R (IN REF. 1).
SQ      SEQUENCE 492 AA; 53906 MW; 58B4ED0D09DC98BB CRC64;

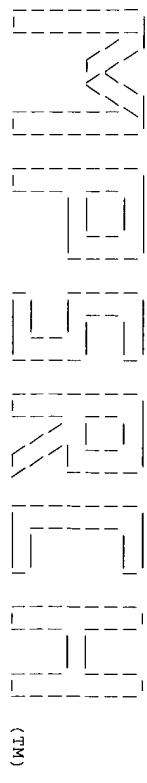
Query Match 2.8%; Score 105; DB 1; Length 492;
Best Local Similarity 33.3%; Pred. No. 7.78e-01;
Matches 16; Conservative 19; Mismatches 9; Indels 4; Gaps 4;

Db      321 TVSLFVVERAGRTLIH-LIGLAGAGCAVMTIALALTEPLPMWSYL 367
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :
Cc      460 TVHSFLVGRVDHAKKDIISLT-ITVGSLL-FRLALL-KVAMLSFL 504
      : 1 1 : : : : : 1 1 : : : : : 1 1 : : : : : 1 1 :

RESULT 15
ID      GTRL_BOVIN      STANDARD:      PRT:      492 AA.
AC      P27674;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      GLUCOSE TRANSPORTER TYPE 1, ERYTHROCYTE/BRAIN.
GN      SLC2A1 OR GLUT1 OR GLUT-1.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Bosdo R.U., Partridge W.M.;
RT      "Molecular cloning of the bovine blood-brain barrier glucose
RT      transporter cDNA and demonstration of phylogenetic conservation of the
RT      5' untranslated region."
RL      Mol. Cell. Neurosci. 1:224-232(1991).

```





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch-pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sun Sep 3 12:36:25 2000; Maspar time 26.23 Seconds  
913.484 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-332-522B-4  
Description: (1-508) from US09332522B.pep  
Perfect Score: 3806  
Sequence: 1 MDPFVFIVLASLVGLYFF.....LLFATALLKVMLSFLRPLL 508

Scoring table: PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p164  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 50.692; Variance 112.058; scale 0.452

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	222	5.8	368	2	T05339	hypothetical protein 3.88e-16
2	221	5.8	383	2	G72777	probable S2P metalloprotease 5.53e-16
3	212	5.6	377	2	A71450	hypothetical protein 1.32e-14
4	206	5.4	378	2	H75213	serine proteinase htr 1.07e-13
5	193	5.1	389	2	H69048	conserved hypothetical protein 9.35e-12
6	175	4.6	386	2	S75400	hypothetical protein 3.88e-09
7	146	3.8	501	2	A69415	conserved hypothetical protein 3.92e-05
8	134	3.5	165	2	D72085	hypothetical protein 1.42e-03
9	125	3.3	4199	2	S76412	hypothetical protein 1.88e-02
10	120	3.2	344	1	T12335	MADH dehydrogenase (u 1.72e-01
11	117	3.1	396	1	C64907	chloramphenicol resist 7.58e-02
12	116	3.0	161	2	T08294	hypothetical protein 2.25e-01
13	113	3.0	207	2	E59124	cobalamin biosynthesis 5.00e-01
14	115	3.0	422	2	H64489	hypothetical protein 2.94e-01
15	115	3.0	447	2	T02547	hypothetical protein 2.94e-01
16	115	3.0	450	1	A64148	hypothetical protein 2.94e-01
17	115	3.0	460	2	JC2194	vasoactive intestinal 2.94e-01
18	115	3.0	495	2	JC2195	vasoactive intestinal 2.94e-01
19	111	2.9	320	2	A46152	A3 adenosine receptor 8.46e-01
20	110	2.9	360	2	S36750	cannabinoid receptor 1.10e+00
21	112	2.9	487	2	E70508	hypothetical protein 6.51e-01
22	109	2.9	641	2	S63545	MADH dehydrogenase (u 1.42e+00
23	108	2.8	138	2	S17941	xpsi protein - Xantho 1.84e+00

24	105	2.8	138	2	T12060	xpsi - Xanthomonas ca	3.93e+00
25	107	2.8	232	2	C72609	hypothetical protein	2.37e+00
26	108	2.8	266	2	I49059	Ly49c - mouse	1.84e+00
27	106	2.8	268	2	A70417	hypothetical protein	3.05e+00
28	106	2.8	285	2	D71105	hypothetical protein	3.05e+00
29	106	2.8	344	2	T12247	NADh dehydrogenase (u	3.05e+00
30	105	2.8	344	2	T12337	NADh dehydrogenase (u	3.93e+00
31	106	2.8	348	2	T12241	NADh dehydrogenase (u	3.93e+00
32	108	2.8	373	2	D71142	hypothetical protein	1.84e+00
33	107	2.8	459	2	JH0594	vasoactive intestinal	2.37e+00
34	105	2.8	492	2	S09705	glucose transport pro	3.93e+00
35	105	2.8	492	2	I45902	glucose transporter t	3.93e+00
36	105	2.8	614	2	B71551	probable s/t protein	3.93e+00
37	108	2.8	764	2	A35956	thyrotropin receptor	1.84e+00
38	106	2.8	764	1	QRHURH	thyrotropin receptor	3.05e+00
39	107	2.8	770	2	S75042	hypothetical protein	2.37e+00
40	103	2.7	186	2	B29835	trps protein - Escher	6.46e+00
41	104	2.7	266	2	I49114	Ly49h - mouse	5.04e+00
42	104	2.7	344	2	T12336	NADh dehydrogenase (u	5.04e+00
43	104	2.7	379	2	S73517	phosphotransferase sy	5.04e+00
44	104	2.7	508	2	T05156	probable glucose tran	5.04e+00
45	103	2.7	600	2	T11889	NADh dehydrogenase su	6.46e+00

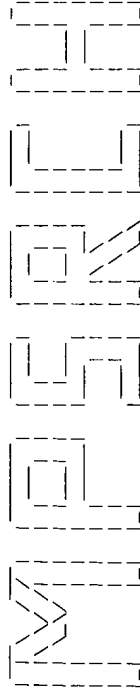
RESULT	ENTRY	TITLE	ORGANISM	DATE	ALIGNMENTS
1	T05339	#type complete	hypothetical protein F1C12.220 - Arabidopsis thaliana	23-Apr-1999	#sequence_revision 23-Apr-1999 #text_change
ACCESSIONS	T05339	215408	Bevan, M.; Terryn, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De Keyser, A.; Neyt, P.; Pouze, P.; Van Den Daele, H.; Villaroel, R.; Gielens, J.; Van Montagu, M.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.	submitted to the Protein Sequence Database, April 1998	
REFERENCE	#accession	T05339	#molecule_type DNA	1-368	#label BEV
#authors	##residues	1-368	##cross-references EMBL:AL022224	##experimental_source	cultivar Columbia; BAC clone F1C12
GENETICS	#map_position	4	104/3; 135/2; 193/3; 220/3		
SUMMARY	#note	F1C12.220	#length 368	#molecular_weight 40315	#checksum 3588
Query Match	Best local Similarity	5.8%;	Score 222;	DB 2;	length 368;
Matches	48;	Conservative	34;	Mismatches	46;
Indels	8;	Gaps	8;		
Db	53	FS-GVAKWMSIGCGFVASILVTVFLLDFHNPPLFSNLTLSAV-rgfs-pstrvsls	109		
Qy	78	FNWGVLTFTSL-DPGLI-LLIATIFSSGSDSSSVSGVGVOLLELLPVDNPLE	134		
Db	110	GIAYLVSTVTVSVHGHLLAAASGIDMEYIAV-FIAIFPGVLAFNDVLOSIPS	168		
Qy	135	EIGYITTVLVCVAVHMGHLLAAVMDVPTVGIGIFIFCL-PLATYELSHDHLNLSRW	193		
Db	169	FNALRYCAGIMHNAV	184		
Qy	194	FRKLRYLCAGIMHNAV	209		
RESULT	2	G72777	#type complete		
ENTRY	probable S2P metalloproteinase	AFE0209	- Aeropyrum pernix		
TITLE	(strain K1)				

ORGANISM	#formal_name	Aeropyrum pernix
DATE	20-Aug-1999	#sequence_revision 20-Aug-1999
ACCESSIONS	G727177	
REFERENCE	A72450	
authors	Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, T.; Kushiida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.	
#journal	DNA Res. (1999) 6:83-101	
#title	Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.	
#cross-references	MUID:99310339	
#accession	G727177	
#status	Preliminary	
#molecule_type	DNA	
#residues	1-383	#label KAW
#cross-references	DBJ:AP000058; NID:65103388; PIDN:BA479121.1; PID:d1042897; PID:g5103600	
#experimental_source	strain K1	
GENETICS		
#gene	AP00209	
SUMMARY	#length 383	#molecular-weight 39729 #checksum 8340
Query Match	5.88	Score 221. DB 2; Length 383;
Best Local Similarity	35.08	Pred. No. 5.53e-16;
Matches	36; Conservative	27; Mismatches 39; Indels 1; Gaps 1;
Db	109 LIPGVTIPMEDLVVAALGVVAHELGHAVVAVEGIRKNAGIALILFIPAFVELD 168	
Qy	125 LLEPVNLEPLEIGYITTVLCVHEHMGALAAVVEDPYGFGRKFCPLPYATELS 184	
Db	169 EEQIMKARLVSRKVESAGVTANILIALLT-LLIANTAPVAEP 210	
Qy	185 HDHLNLSRWFKRRLVLCAGIMHNFVEAGVYLLISTVGITMSP 227	
RESULT	3	
ENTRY	A71450	#type complete
TITLE	hypothetical protein PH0256 - Pyrococcus horikoshii	
ORGANISM	#formal_name Pyrococcus horikoshii	
DATE	14-Aug-1998	#sequence_revision 14-Aug-1998
ACCESSIONS	A71450	
REFERENCE	A71000	
authors	Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ofuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, T.; Kushiida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuma, H.; Kikuchi, H.	
#journal	DNA Res. (1998) 5:55-76	
#title	Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3	
#cross-references	MUID:98344137	
#accession	A71450	
#status	Preliminary; nucleic acid sequence not shown; translation not shown	
#molecule_type	DNA	
#residues	1-377	#label KAW
#cross-references	GB:AP000001; NID:93236128; PID:d1030271; PID:g3256645	
#experimental_source	strain OT3	
#note	this accession replaces an interim accession for a sequence replaced by Genbank	
GENETICS		
#gene	PH0256	
SUMMARY	#length 377	#molecular-weight 41845 #checksum 1083

	Query Match	5.6%	Score 212;	DB 2;	Length 377;	
	Best Local Similarity	32.7%	Pred. No.	1.32e-14;	Indels 4; Gaps 4	
	Matches	37;	Conservative	35;	Mismatches 37;	Indels 4; Gaps 4
Dd	97 PTMPS-VOLVIVPTPIPLV-YGL-IALAI-P-VIWHESHSFVARSEDIPKSGLLLEFI 152    :   ::::     :     :   :   ::::     :     :   :					
Oy	115 PVGVYQVEILLPQVNPLEIEIGYTITVLCLVHEHMGHALAAMDEVVTGFGICIFIF 174 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Dd	153 ITPGAWEPPDEDLKKAPLRSLRLEVFGAGSFANFVALISILLVNGIALAFEP 205 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Oy	175 CLPIAYTELSDHLNLSRMFRKLRYLCAGIMHNFFVGVCYLSTVGTMTSP 227 : :   :   :   :   :   :   :   :   :   :   :   :   :					
RESULT ENTRY TITLE ORGANISM DATE	4 H75213 #type complete serine proteinase hira related protein PAB2195 - Pyrococcus abyssi (strain Orsay) #formal_name Pyrococcus abyssi 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999					
ACCESSIONS REFERENCE	H75213 A75001 anonymous, Genoscope submitted to the EMBL Data Library, July 1999 Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution. #accession H75213					
#authors	##status Preliminary ##molecule_type DNA ##residues 1-378 ##label KAW ##cross-references GB:AJ24883; GI:AO96836; NID:g5457433; PID:CABA9159.1; PID:e1515053; FID:g5457668					
GENETICS SUMMARY	#experimental_source strain Orsay PAB2195 #length 378 #molecular_weight 41715 #checksum 4202					
Query Match	5.4%; Score 206; DB 2; Length 378; Best Local Similarity 30.2%; Pred. No. 1.07e-13; Matches 49; Conservative 44; Mismatches 58; Indels 11; Gaps 9;					
Dd	101 VPS-VOLVIVPTPIPLV-YGL-ISLAU-LIIWELSHGFVAARENPDKSGLLFLVP 156    :   ::::     :     :   :   ::::     :     :   :					
Oy	118 VPVOEELLRPVNLPLEIEIGYTITVLCLVHEHMGHALAAVEDVVTGFGICIFICLP 177 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Dd	157 GAWEPEDEDELKRAPLRSLRVFAAGSFANFYAIS-VIVEN-GVTLA--FE-PRGEV 211 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Oy	178 LAYTELSHDHLNSLRMRKRLRYLCAGIMHNFFVGVCYLSTVGTMTSPLEYAVNQVVV 237 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Dd	212 FGVIKDSPAIGI--LEKGDIYVELINGRKINTLEEFTFMNTT 251 : :   :   :   :   :   :   :   :   :   :   :   :   :					
Oy	238 TELTRKSPLRGERGLQVDNOITYNGCPNVSESWATCLONS 279 : :   :   :   :   :   :   :   :   :   :   :   :   :					
RESULT ENTRY TITLE ORGANISM DATE ACCESSIONS REFERENCE	5 H69048 #type complete conserved hypothetical protein MTH1368 - Methanobacterium thermoautotrophicum (Strain Delta H) #formal_name Methanobacterium thermoautotrophicum 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998					
#authors	H69048 A69000 Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldrege, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumma, W.; Potlauer, B.; Qiu, D.; Spadafora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanji, N.; Cartuso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrkovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.					
#journal	J. Bacteriol. (1997) 179:7135-7155					



\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:37:38 2000; MasPar time 14.63 Seconds  
Tabular output not generated. 822.382 Million cell updates/sec

Title: >US-09-332-522B-4  
Description: (1-508) from US09332522B.pep  
Perfect Score: 3806  
Sequence: 1 MDPVFVFIVLASLYGVLYFF.....LLFALLKLVAMLSFLRPLL 508

Scoring table: PAM 150  
Gap 11

Searched: 189963 seqs, 23696106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 36.607; Variance 170.723; scale 0.214

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	2.9	763	1 R11341	Human Thyroid Stimulat	8.63e+00
2	108	2.8	320	1 R79450	Rat A3 adenosine recep	1.37e+01
3	107	2.8	360	1 R67001	Cannabinoid receptor H	1.60e+01
4	107	2.8	459	1 R42848	VIP receptor protein.	1.60e+01
5	106	2.8	674	1 R12917	Human TSH Receptor wit	1.87e+01
6	106	2.8	714	1 R12915	Human TSH Receptor wit	1.87e+01
7	106	2.8	756	1 R12916	Human TSH Receptor wit	1.87e+01
8	106	2.8	756	1 R24244	Rat thyrotropin recept	1.37e+01
9	106	2.8	764	1 R12914	Human Thyroid Stimulat	1.87e+01
10	106	2.8	764	1 R12505	Human Thyroid Stimulat	1.87e+01
11	106	2.8	764	1 R13269	Human Thyroid Stimulat	1.87e+01
12	101	2.7	314	1 R27874	Odorant receptor clone	3.99e+01
13	103	2.7	336	1 R48733	G-protein coupled huma	2.95e+01
14	103	2.7	336	1 W02705	G-protein coupled huma	2.95e+01
15	103	2.7	428	1 W56671	Caenorhabditis elegans	2.95e+01
16	101	2.7	492	1 R11360	Glucose Transporter Pr	3.99e+01
17	98	2.6	216	1 R75910	Human olfactory recept	6.26e+01
18	98	2.6	216	1 R75909	Human olfactory recept	6.26e+01
19	99	2.6	312	1 R27873	Odorant receptor clone	5.39e+01
20	99	2.6	314	1 R27876	Odorant receptor clone	5.39e+01
21	98	2.6	365	1 W20791	H. pylori transmembran	6.26e+01
22	100	2.6	407	1 R04572	ORF4 product from the	4.64e+01
23	98	2.6	457	1 W69556	Rat equilibrative nucl	6.26e+01

24	100	2.6	492	1 W17835	Human glucose transpor	4.64e+01
25	99	2.6	492	1 W30000	Human GLUT1 protein.	5.39e+01
26	99	2.6	762	1 W55039	Homo sapiens TSH recep	5.39e+01
27	99	2.6	762	1 W14778	Human TSH receptor.	5.39e+01
28	98	2.6	2100	1 W89579	Calcium permeable volt	6.26e+01
29	98	2.6	2104	1 W57773	Musca domestica voltag	6.26e+01
30	98	2.6	2104	1 W89578	Calcium permeable volt	6.26e+01
31	98	2.6	2105	1 W89577	Calcium permeable volt	6.26e+01
32	98	2.6	2105	1 W57772	Musca domestica voltag	6.26e+01
33	97	2.5	263	1 R25424	Human MIP.	7.26e+01
34	95	2.5	284	1 W02718	G-protein coupled odor	9.75e+01
35	96	2.5	303	1 W32448	Mycobacterium tubercul	8.41e+01
36	96	2.5	303	1 W32380	Mycobacterium tubercul	8.41e+01
37	96	2.5	303	1 W81682	M. tuberculosis immuno	8.41e+01
38	96	2.5	303	1 W54320	Mycobacterium tubercul	8.41e+01
39	95	2.5	342	1 W81576	EBV-induced G-protein	9.75e+01
40	95	2.5	359	1 Y00890	Canine histamine H2 re	9.75e+01
41	95	2.5	365	1 W23606	Human P2Y4 receptor po	9.75e+01
42	96	2.5	514	1 R95044	Apoptosis participatin	8.41e+01
43	97	2.5	574	1 W81559	Respiratory syncytial	7.26e+01
44	97	2.5	574	1 R39285	Respiratory syncytial	7.26e+01
45	95	2.5	574	1 W47604	HRSV glycoprotein F.	9.75e+01

ALIGNMENTS

RESULT 1  
ID R11341 standard; Protein; 763 AA.  
AC R11341;  
DT 05-JUN-1991 (first entry)  
DE Human Thyroid Stimulating Hormone receptor.  
KW thyroid stimulating hormone; TSH; receptor; Graves' Disease;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT peptide 1..34  
FT domain /label= signal peptide  
FT domain 35..418  
FT domain /label= extracellular domain  
FT domain 419..440  
FT domain /label= transmembrane domain I  
FT domain 441..471  
FT domain /label= transmembrane domain II  
FT domain 495..517  
FT domain /label= transmembrane domain III  
FT domain 538..560  
FT domain /label= transmembrane domain IV  
FT domain 581..602  
FT domain /label= transmembrane domain V  
FT domain 625..648  
FT domain /label= transmembrane domain VI  
FT domain 660..681  
FT domain /label= transmembrane domain VII

WO9103483-A.  
21-MAR-1991.  
07-SEP-1990; U05066.  
08-SEP-1989; US-404899.  
10-AUG-1990; US-565669.  
(NEWB-) NEW ENGLAND MED CEN.  
Cone R;  
WPI; 91-102023/14.  
N-PSDB; Q11124.  
Pure nucleic acid encoding thyroid stimulating hormone and protein - used to detect increased levels of intracellular C-AMP indicative of thyroid malfunction, and to treat Graves disease and thyroid conditions  
Claim 7; Fig 6; 3pp; English.  
Total RNA was isolated from a human Graves' thyroid and a bovine thyroid sample and separately reverse transcribed to give cDNA. Five micro litres of the cDNA was used in a PCR with two degenerate oligonucleotides having homology to transmembrane domains III and IV of human Luteinizing Hormone receptor (see Q1121 and Q1122). The amplified DNA was isolated, digested with HindIII and EcoRI and

CC subcloned in vector pBS- for transformation into E.coli. The  
 CC fragments were removed from the vector, labelled and used as probes  
 CC to screen a lambda gtl library constructed from RNA from the  
 CC thyroid of a Graves' patient. One strongly hybridising clone,  
 TR.12.6-1 (HTSH receptor), was sequenced and found to contain a  
 CC full-length human fSH receptor cDNA. The protein sequence it encodes  
 CC was deduced from the nucleotide sequence.  
 CC See also Q1117-Q1123.  
 SQ Sequence 763 AA;

Query Match 2.9%; Score 111; DB 1; Length 763;

Best Local Similarity 25.0%; Pred. No. 8.63e+00; Indels 12; Gaps 12;  
 Matches 40; Conservative 40; Mismatches 68; Mismatches 12;

Db 434 FVLILLTSHYK-LNV-PRFLM-C-NLAFAFCMGVYLLIASVDLYTHSEYVNHAI-DW 488

QY 4 FVFIVLASLYGVLYFFDFRFSKSMHPYDAF-LKNTGLSINFMSLHWHTSAFNRTLLRW 62

Db 489 -QTGPGGNTAGFTTFA-SLSVYTLVITLERWYATFAMALDKRLRHACAIMYGGW 546

QY 63 GSAGNSCTRRVMTSFNVGLVTFSLPIGLILLIATIFSSG-EQDSSSVSPVGPVQ 121

Db 547 VCCFLAL-LPLVGISSY-AKVSICLPMDTETPLALAVIV 584

QY 122 LEILLPGVNLPLEIGYITTLVLCVVH-EMGHAAVM 160

RESULT 2  
 ID R79450 standard; Protein; 320 AA.

AC R79450;

DT 19-JAN-1996 (first entry)

DE Rat A3 adenosine receptor.

KW A3 adenosine receptor; screening; identification; agonist;

KW antagonist; drug; probe; detection; genetic disorder; RFLP;

KW restriction fragment length polymorphism.

OS Rattus rattus.

PN US5441883-A.

PF 15-AUG-1995.

PR 03-MAR-1992; 847563.

PR 03-MAR-1992; US-847563.

PR 02-AUG-1993; US-101435.

PA (UYOR-) UNIV OREGON HEALTH SCI.

PI Civelli O, Zhou Q.

DR WPI; 95-292492/38.

DR N-PSDB; Q97312.

PT DNA and protein sequence(s) for novel adenosine receptor, A3 - also

PT recombinant vector contg the DNA, DNA probe(s), and cell culture

PT transformed with the vector.

PS Claim 1; Figure 1; 29pp; English.

CC A3 adenosine receptors produced from cloned genes may be used to

CC screen compounds for A3 adenosine receptor activity, or for

CC determining the amount of adenosine agonist or antagonist drug in a

CC solution. The DNA or its fragments may also be used as probes to

CC determine tissue distribution of the receptors, to detect the

CC presence or absence of the gene or in RFLP to detect genetic

CC disorders.

SQ Sequence 320 AA;

Query Match 2.8%; Score 108; DB 1; Length 320;

Best Local Similarity 26.5%; Pred. No. 1.37e+01;

Matches 31; Conservative 35; Mismatches 39; Indels 12; Gaps 10;

Db 51 YFIVSLADIAVGLVPIPLPIAVSLEVMHFYACLFMSCVLLVFTHASIMSLAIADVR 110

QY 138 YIITLVLC-LVVHEMGHALA-AWVEDVPTVGFGIKFICLPLAYTELH-DHLN-SL-R 192

Db 111 YLRVRLTVRYRTVTQRRIWFLGLCWL-VSFL-VGLTFMGWNRKVTL-ELSONSS 164

QY 193 WFR-KLRVLCAGI---WHNFVAGVCYLLISTVGTITNSPLVAYNQHVVVVTELTRKSP 245

RESULT 3

ID R67001 standard; Protein; 360 AA.

AC R67001;  
 DT 24-JUN-1995 (first entry)  
 DE Cannabinoid receptor HCX5.36.  
 KW Cannabinoid receptor; CX5.36; cannabinoid screening.  
 OS Homo sapiens.  
 PN WO9425589-A.  
 PD 10-NOV-1994.  
 PF 22-APR-1994; G00863.  
 PR 27-APR-1993; GB-008679.  
 PR 21-OCT-1993; GB-021697.  
 PA (MEDI-) MEDICAL RES COUNCIL.  
 PI Munro BS; 94-358267/44.  
 DR P-PSDB; Q74797.

DR New nucleic acid encoding non-brain cannabinoid receptor - and  
 PT related vectors, transformed cells and polypeptide(s), useful for  
 PT screening cannabinoid cpds. with reduced psychoactive effects.  
 PS Claim 3; Fig 1; 4pp; English.  
 CC Cannabinoid receptor CX5.36 is not normally expressed in mammalian  
 CC brain tissue. CX5.36 is used to screen for compounds with  
 CC cannabinoid activity but reduced psychoactive effects which are  
 CC useful as analgesic, antiinflammatory, immunosuppressant,  
 CC anticonvulsant, antiglaucoma and antiemetic agents.  
 SQ Sequence 360 AA;

Query Match 2.8%; Score 107; DB 1; Length 360;

Best Local Similarity 23.5%; Pred. No. 1.60e+01;

Matches 39; Conservative 49; Mismatches 64; Indels 14; Gaps 12;

Db 129 IDRYL--CLRYPPSYKALLTR-GRALVTGLIMVLSALVYLPMLGMWTCPCRCSELFPL 185

QY 20 FDRFFKSCMHP--YDAFLKNTGLSINFMSLHWHTSAFNRTL-LR-WGSAGNSCTRRV-M 74

Db 186 IPNDYLLSWLFTAFISGIIYVGHVLMKAHQHVAS-LSGHODRQVPGMARVLDVRLA 244

QY 75 ITS-FNVGLVTFSLPIGLILLIATIFSSGEQDSSSVSPVGPVQ-LEILLPGVNL 132

Db 245 -KTIGLVAVLLICWFPVLPALMAHSLATTLSDQVKAFACSMCLC 289

QY 133 LEEIGYITTLVLC-LVWHE-MGHAAVMDVPTVGFGIKFICL 176

RESULT 4

ID R42848 standard; Protein; 459 AA.

AC R42848;

DT 13-MAY-1994 (first entry)

DE VIP receptor protein.

KW vasoactive intestinal polypeptide receptor; VIP; rat; binding;

KW adenylylate cyclase activity; stimulus.

OS Rattus rattus.

PN J05255394-A.

PD 05-OCT-1993.

PF 13-FEB-1992; 026607.

PR 13-FEB-1992; JP-026607.

PA (OSAB-) ZH OSAKA BIOSCIENCE KENKYUSHO.

DR WPI; 93-348480/44.

DR N-PSDB; Q50349.

PT Vasoactive intestinal polypeptide - prepd. in large amt. by

PT culturing microbe transformed by new DNA coding polypeptide

PS Claim 2; Page 6; 14pp; Japanese.

CC The sequence can be used to produce large amounts of the VIP

CC receptor peptide, by culturing a microorganism transformed by

CC the sequence.

SQ Sequence 459 AA;

Query Match 2.8%; Score 107; DB 1; Length 459;

Best Local Similarity 22.4%; Pred. No. 1.60e+01;

Matches 30; Conservative 40; Mismatches 54; Indels 10; Gaps 9;

Db 228 MANF-FWLLVEGLY--LYTLLAVSFFSERKYF-WGYTLICGWVSPVFTTI-WTVVRIYE 282

QY 1 MDPFVFFIVLASLYGVLY-EFD-RFFKSCMHYPYDAFLKNTGLSINFMSLHWHT-SAFNR 57

```

#title      Complete genome sequence of Methanobacterium
            thermoautotrophicum Delta H: functional analysis and
            comparative genomics.
#cross-references MUID:98037514
#accession  H69048
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
##residues 1-389 ##label MTH
##cross-references GB:AF000899; GB:AF000666; NID:g2622468; PID:g2622474
##experimental_source strain Delta H
GENETICS
#gene       MTH1368
#start_codon GTG
SUMMARY
#length 389 #molecular-weight 42956 #checksum 2072
Query Match 5.1%; Score 193; DB 2; Length 389;
Best Local Similarity 29.5%; Pred.No. 9.35e-12;
Matches 28; Conservative 34; Mismatches 32; Indels 1; Gaps 1;

Db 113 LGXGIIGLATVIVVHFAHGALRGVRIKSGLLLAILPCGAFVPEDEEDIKIRPIS 172
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 136 IGYIITLVLCVHENGHALAAMEDVPTGTGKIFCFLPLAYTELSHDHLNLSLRWR 195
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

Db 173 KMFRIYAGSVANILAGICFALFFGSSFPAMPAF 207
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 196 KLRVLCAGIWHNFVAGVCY-LLISTVGITMSPLY 229
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

RESULT 6
ENTRY S75400 #type complete
TITLE hypothetical protein c04034 - Sulfolobus solfataricus
ORGANISM "formal_name Sulfolobus solfataricus"
DATE 11-Oct-1997 #sequence_revision 24-Oct-1997 #text_change
      17-Mar-1999
ACCESSIONS S75400
REFERENCE S73076
#authors  Sensesen, C.W.; Klenk, R.K.; Singh, R.K.; Allard, G.; Chan,
            C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.;
            Gaasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois,
            R.L.
#journal  Mol. Microbiol. (1996) 22:175-191
#title    Organizational characteristics and information content of an
            archaeal genome: 156 kb of sequence from Sulfolobus
            solfataricus p2.
#cross-references MUID:97055432
#accession S75400
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-386 ##label SEN
##cross-references EMBL:Y08257; NID:g1707772; PID:e283840; PID:g1707806
##experimental_source strain p2
#note      the nucleotide sequence was submitted to the EMBL Data
            Library, September 1996
GENETICS
#start_codon TTG
SUMMARY
#length 386 #molecular-weight 43423 #checksum 5980
Query Match 4.6%; Score 175; DB 2; Length 386;
Best Local Similarity 26.0%; Pred.No. 3.88e-09;
Matches 47; Conservative 51; Mismatches 78; Indels 5; Gaps 5;

Db 87 LPIGFLMIAGIVTILYIIEMTLTKPNOTPTTALAPIPGVITISQLPYILLATGVSV 146
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 89 LPICGLIILNATIFSSEQDSSSVSPGVVP-VOLIEILLPGVNLPEEIGYITTLVLC 147
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

Db 147 AIHEIFHALSATSNVKNVKNVGLLIGIFPGAFVPEDEDFNKSTSNAKLIIAAGIVIN 206
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 148 VVHEMGHALAAMVEDVPTGTGKIFCFLPLAYTELSHDHLNLSRWRKLRVLCAGIWHN 207
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

Db 207 LVIALIA-LPLS-FEILDYP-SALSOGIIIEGVNLTNP-ANASLHTGDIIVSYNGRIT 262
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 208 FVAGVCYLLISTVGTIMTSPLYAYNOHVITELTRKSPLRGERGLVDNOITOVNGCPVN 267
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

```

[illegible]

```

Query Match          3.1%; Score 117; DB 1; Length 396;
Best Local Similarity 26.4%; Pred. No. 1.72e-01;
Matches 33; Conservative 42; Mismatches 41; Indels 9; Gaps 8;

Db 239 IAGSFANFATALLLGGAGIGSVFKIGNOVASALVSTATAALLVLCIALLLPAAANSE 298
QY 75 ITSFNVGLVTFSLPLGLILLIATIFSS-GEQDSSSVSPVGPV-VQLEILLPGVN-- 130

Db 299 IHGLVLSIFWGIAMITGLGMQVKVLAAPDAIDVAMALFS-G-IFNIGIGAGALVGNQV 356
QY 131 LPLEEIG-VY-IITLVLCVLVHMGHALAAMVEDVPTVGFKIFCLPLAYTELSDHL 188

Db 357 -SLHW 360
QY 189 NSLRW 193

RESULT 12
ENTRY T08294 #type complete
TITLE hypothetical protein H0875 - Halobacterium sp. (strain NRC-1)
ORGANISM plasmid pNRC100
#variety strain NRC-1
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
ACCESSION T08294
REFERENCE Z16408
#authors Ng, W.V.; Ciuffo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall, B.; Loretz, C.; Seto, J.; Slagel, J.; Hood, L.; Dassarma, S.

#journal Genome Res. (1998) 8:1131-1141
#title Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or minichromosome?
#cross-references EMBL:AF016485; NID:g2822278; PID:g2822355
#experimental_source strain NRC-1
GENETICS
#genome plasmid pNRC100
SUMMARY #length 161 #molecular-weight 16413 #checksum 5265

Query Match          3.0%; Score 116; DB 2; Length 161;
Best Local Similarity 22.0%; Pred. No. 2.25e-01;
Matches 22; Conservative 28; Mismatches 47; Indels 3; Gaps 3;

Db 44 FVLSMGLAPLVAVGPAQAOSQVDVCDTGATGDLVFGAVAGLGLPATGYTGKAG 103
QY 86 FSLPLGLILLIATIFSSGQDSSSVSPVGPVQLEILLPGVN-LPLEEIGYITLV 144

Db 104 LSYWRAGNPEKNDAKEKLVMSCIGFGIV-TIALVSPEL 142
QY 145 LC-LVVHMGHALAAMVEDVPTVGFKIFCLPLAYTEL 183

RESULT 13
ENTRY E69124 #type complete
TITLE cobalamin biosynthesis protein M related protein - Methanobacterium thermoautotrophicum (strain Delta H)
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSION E69124
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.
#cross-references MUID:98037514
#accession E69124
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-207 #label MTH
#cross-references GB:AE000807; GB:AE000566; NID:g2621239; PID:g2621245
#experimental_source strain Delta H
GENETICS
#gene MTH200
#start_codon GTG
#classification #superfamily cobalamin biosynthesis protein M
SUMMARY #length 207 #molecular-weight 22151 #checksum 3365

Query Match          3.0%; Score 113; DB 2; Length 207;
Best Local Similarity 31.6%; Pred. No. 5.00e-01;
Matches 25; Conservative 23; Mismatches 26; Indels 5; Gaps 5;

Db 19 VALINIGIFYIFSKPKERRIVIGLFAAAAASVSPSPGVPVHF-FLIPLAAI 77
QY 75 ITSFNVGLV-TFSLPLGILLIAT-IFSSGQDSSSVSPVGPVQLEILLPGVNL 131

Db 78 LLGPLTAVIVA-TLCIVIQ 95
QY 132 PLEEIGYITLVLCVLVH 150

RESULT 14
ENTRY H64489 #type complete
TITLE hypothetical protein MJ1521 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
ACCESSION H64489
REFERENCE A64300
#authors Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kierlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.L.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession H64489
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-422 #label BUL
#cross-references GB:U67593; GB:L77117; NID:gi592149; PID:gi500411; TIGR:MJ1521
GENETICS
#map_position FOR1498907-1500175
#start_codon GTG
#classification #superfamily hypothetical protein MJ0057
SUMMARY #length 422 #molecular-weight 45804 #checksum 5126

Query Match          3.0%; Score 115; DB 2; Length 422;
Best Local Similarity 23.9%; Pred. No. 2.94e-01;
Matches 22; Conservative 32; Mismatches 33; Indels 5; Gaps 5;

```

[illegible]

RESULT	15
ENTRY	T02547
TITLE	hypothetical protein T6B15.4 - Arabidopsis thaliana
ORGANISM	#formal_name Arabidopsis thaliana #common_name mouse-ear Cress
DATE	05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 23-Apr-1999

Accession number	Reference	Authors	Submission date	Description
T02347	214678	Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.		submitted to the EMBL Data Library, July 1998
				Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.

```
#accession      T02547
##status        translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues      1-447 ##label ROU
##cross-references EMBL:AC004681; NID:g3298532; PID:g3298536
##experimental_source cultivar Columbia
```

CONFIDENTIAL
CONFIDENTIAL SOURCE CULTURAL COLUMBIA
GENETICS
#map_position 2
#note T26B15.4
SUMMARY #length 447 #molecular-weight 48726 #checksum 89

```

SUMMARY
#length 447 #molecular-weight 48726 #checksum 8894

Query Match 3.0% Score 115; DB 2; Length 447;
Best Local Similarity 37.5%; Pred. No. 2,94e-01;
Matches 18; Conservative 14; Mismatches 15; Indels 1; Gaps 1;

Db 77 ISGIDLGSFESVLEALAVLTITIIWHESGHRFLAASLQGIHVSKFAIGF 124
QV 126 LKPVNL-PLKEIGYITITVLGVHMGHALAAVMDVPVTGFGIKF 172

```

Search completed: Sun Sep 3 12:37:21 2000  
Job time : 56 secs.



RESULT	8	standard: Protein; 764 AA.
ID	R24244;	
AC	R24244;	
DT	01-DEC-1992	(first entry)
DE	Rat thyrotropin receptor precursor.	
KW	Thyroid dysfunction; TSH; thyroid stimulating hormone;	
KM	Thyroid gland; FRTL-5 thyroid epithelial cell line; strain Fischer	
KW	transmembrane domain; Graves' disease; hyperthyroidism.	
OS	Rattus norvegicus.	
FH	Key	Location/Qualifiers
FT	peptide	1..21
FT	/label= signal_peptide	
FT	protein	22..764
FT	/label= mature_TSH_receptor	
FT	modified_site	77..79
FT	/label= N-glycosylation	
FT	modified_site	99..101
FT	/label= N-glycosylation	
FT	modified_site	177..179
FT	/label= N-glycosylation	
FT	modified_site	198..200
FT	/label= N-glycosylation	
FT	modified_site	302..304
FT	/label= N-glycosylation	
FT	domain	1..400
FT	/label= extracellular	
FT	domain	401..681
FT	/label= transmembrane	
FT	domain	416..440
FT	/label= hydrophobic	
FT	/note= "TM1"	
FT	domain	451..473
FT	/label= hydrophobic	
FT	/note= "TM2"	
FT	domain	495..516
FT	/label= hydrophobic	
FT	/note= "TM3"	
FT	domain	538..560
FT	/label= hydrophobic	
FT	/note= "TM4"	
FT	domain	582..605
FT	/label= hydrophobic	
FT	/note= "TM5"	
FT	domain	626..648
FT	/label= hydrophobic	
FT	/note= "TM6"	
FT	domain	662..681
FT	/label= hydrophobic	
FT	/note= "TM7"	
FT	modified_site	748..750
FT	/label= phosphorylation_site	
FT	/note= "potential"	
N	WO9208726-A.	
PD	29-MAY-1992.	
PE	15-NOV-1990:	U06533.
PR	15-NOV-1990:	WO-U065339.
PA	(AKAM/) AKAMIZU T.	
PA	(BANT/) BAN T.	
PA	(COHN/) COHN L D.	
PA	(IKUY/) IKUYAMA S.	
PA	(KOSU/) KOSUGI S.	
PA	(SAJI/) SAJI M.	
P1	Akamizu T, Ban T, Ikuyama S, Kohn LD, Kosugi S, Saji M,	
DR	WP1: 92-200128/24.	
DR	N-PSDB: Q25387.	
PT	Nucleotide expressed from rat thyrotropin receptor gene - and	
PT	corresp. polypeptides and monoclonal antibodies, useful for	
PT	diagnosis and treatment of thyroid disorders, in gene therapy	
P5	Disclosure: Fig 1; 44pp; English.	
CC	The full-length rat thyrotropin receptor sequence was deduced from	
CC	the coding sequence deposited in the Genbank data base under	

Query Match	2.8%	Score 108	DB 1	Length 764
Best Local Similarity	28.0%	Pred. No. 1.37e+01		
Matches	28	Conservative	22	Mismatches 42; Indels 8; Gaps 8
Db	434	FVLFLVLTSHK-LTV-PRFLM-C-NLAFLDFCGVLLILIASYDLTHREYVNHAI-DW	488	
Oy	4	FVFVFLVILASLXGLVLPDFDRFKSCMHPRYDAF-LKNGLSLINFNSLMHNSAFNRILLRW	62	
Db	489	-QTGPGCNFAGFTVFA-SELSTVTLTVITLERNVATIFA	526	
Oy	63	GSAGNSCTRRVMIRTSFNVGLVLPFLSLPIGILLIATIFS	102	
RESULT	9			
ID	R12914	standard; Protein; 764 AA.		
AC	R12914			
DE	25-SEP-1991	(first entry)		
PR	Human Thyroid Stimulating Hormone Receptor.			
TM	TSH; thyrotropin; Grave's disease.			
OS	Homo sapiens.			
FT	domain	Location/Qualifiers		
FT		1..418		
FT		/label= extracellular		
FT		/note= "putative"		
FT	domain	419..682		
FT		/label= transmembrane		
FT		/note= "putative"		
FT	domain	683..764		
FT		/label= short intracellular cytoplasmic		
FT		/note= "putative"		
FT	modified_site	99		
FT		/label= N-glycosylation site		
FT	modified_site	113		
FT		/label= N-glycosylation site		
FT	modified_site	177		
FT		/label= N-glycosylation site		
FT	modified_site	198		
FT		/label= N-glycosylation site		
FT	modified_site	302		
FT		/label= N-glycosylation site		
PN	W09109137-A.			
PD	19-JUN-1991.			
PF	19-DEC-1990; U07387.			
PR	20-DEC-1989; US-451973.			
PR	30-AUG-1990; US-575018.			
PA	(RAPO//) RAPOPORT B.			
PI	Rapoport B:			
DR	WPI: 91-208163/28.			
DR	N-PSDB: Q12529.			
PT	Recombinant, enzymatically active thyrotropin receptor			
PT	generated in non-thyroidal eukaryotic cells, useful for study of			
PT	Graves disease and treatment of thyrotoxicosis			
PS	Claim 3; Fig 1; 160pp: English.			
CC	A human thyroid cDNA library was screened with two synthetic probes			
CC	based on the reported amino acid sequence of the 3rd and 4th			
CC	transmembrane domains of a putative human thyrotropin receptor and			
CC	related receptors. This sequence was deduced from the nucleotide			
CC	sequence of a 4th clone isolated from the library. The protein			
CC	contains a putative signal peptide, 7 transmembrane domains (within			
CC	the general transmembrane domain defined in the features table),			
CC	5 potential glycosylation sites and a very short intracytoplasmic			
CC	region. See also Q12530-1 and R12915-7.			
SO	Sequence 764 AA;			
Query Match	2.8%	Score 106	DB 1	Length 764
Best Local Similarity	25.0%	Pred. No. 1.87e+01		
Matches	40	Conservative	39	Mismatches 69; Indels 12; Gaps 12



D	b	434	FVLLILLTSHRK-LNV-PRFLM-C-NLAFAFCMGMYLLIASVDLYTHSEYVNHAI-DW	488
Q	y	4	FVFHFVLASLVGLVDFEDRFRFSKCHMHPYDAF-LKNTGLSINFSMLHMTSAFNRLLRW	62
D	b	489	-QTGGCCTACGFFVFA-SELSVYTLVITLERWVAIFEARLDRKIRLRACAIMGW	546
Q	y	63	GSAGNCSCTRRMWTISFNVCGLVLTSLDPLIGILLIATITFS-SGQDSSSSVSPVGVPVQ	121
D	b	547	VCCFLAL-LPLVGISY-AKVSICLPMDTTPALAYIV	584
Q	y	122	LEILLPGVNLPLEEIGYITTLVCLVNH-EMGHALAAM	160
RESULT 10				
ID			R12505 standard: Protein: 764 AA.	
AC			R12505:	
DT			10-SEP-1991 (first entry)	
DE			Human thyroid stimulating hormone.	
KW			TSH: thyrotropin; oocyte; anti-TSHr.	
OS			Homo sapiens.	
PN			EP-433509-A.	
PD			26-JUN-1991.	
PF			14-DEC-1989: 403493.	
PR			14-DEC-1989: EP-403493.	
PA			(HENN-) HENNING, BERLIN GMBH.	
PI			Parmentier M, Libert F, Dumont J, Vassart G;	
DR			WPI: 91-186499/26.	
DR			N-PSDB: Q12164.	
PT			Polypeptide with thyrotropin receptor activity - used to	
PT			quantitatively detect thyrotropin or its antibodies.	
PS			Disclosure, Fig 8; 31pp; English.	
CC			The sequence is derived from a human genomic library by screening	
CC			with a canine probe.	
CC			The human TSHr may be used to quantitatively detect TSH or anti-TSH	
CC			recombinant Abs, and labeled Abs raised to the TSHr product may be	
CC			used to detect TSHr.	
SO			Sequence 764 AA:	
Query Match 2.8%; Score 106; DB 1; Length 764;				
Best Local Similarity 25.0%; Pred. No. 1.87e+01;				
Matches 40; Conservative 39; Mismatches 69; Indels 12; Gaps 12;				
D	b	434	FVLLILLTSHRK-LNV-PRFLM-C-NLAFAFCMGMYLLIASVDLYTHSEYVNHAI-DW	488
Q	y	4	FVFHFVLASLVGLVDFEDRFRFSKCHMHPYDAF-LKNTGLSINFSMLHMTSAFNRLLRW	62
D	b	489	-QTGGCCTACGFFVFA-SELSVYTLVITLERWVAIFEARLDRKIRLRACAIMGW	546
Q	y	63	GSAGNCSCTRRMWTISFNVCGLVLTSLDPLIGILLIATITFS-SGQDSSSSVSPVGVPVQ	121
D	b	547	VCCFLAL-LPLVGISY-AKVSICLPMDTTPALAYIV	584
Q	y	122	LEILLPGVNLPLEEIGYITTLVCLVNH-EMGHALAAM	160
RESULT 11				
ID			R13289 standard: Protein: 764 AA.	
AC			R13289:	
DT			15-OCT-1991 (first entry)	
DE			Human Thyroid Stimulating Hormone Receptor.	
KW			TSH: thyrotropin; Ashimoto's Disease; Basedow's Disease.	
OS			Homo sapiens.	
FH			Key	
FT			peptide	
FT			domain	
FT			modified_site	
FT			modified_site	
FT			modified_site	

Query Match	2.8%	Score 106	DB 1	Length 764
Best Local Similarity	25.0%	Pred. No. 1.87e+01		
Matches	40	Conservative	39	Mismatches 69; Indels 12; Gaps 12.
Db	434	FVLLILTSYRK-LNV-PRFLM-C-NLAFDFCMGYLLIASVLYLTSEYVNHAI-DW	488	
Oy	4	FVFVFLVSLGVLGYLFEDRFKSCMHYPDAF-LKNTGLSINFSMLHMHSAEYRLLRW	52	
Db	489	-QTGGCGNTAFAFFVFA-SELSVTLTVITLERYATLTFAMRLDRKRLRHACAIWGW	546	
Oy	63	GSAGNSCRRKRWITSPFNGVLYTSLPIGLILDAIFIS-SEGODSSSVSPVGPVQ	121	
Db	547	VCCFLAL-LPLVAGISY-AKVSICLPMDETPLALAIY	584	
Oy	122	LEILLPGVNLPLEELIGYITITLVCLVYH-EMGHALAAM	160	
RESULT	12			
ID	R27874	standard: Protein; 314 AA.		
AC	R27874			
DE	15-MAR-1993	(first entry)		
KW	Odorant receptor; insect; vertebrate; fish; mammal; neurotransmitter; hormone; G-protein; surface receptor; olfactory epithelium; PCR;			
KW	Sprague-Dawley rat; amplification; primer; polymerase chain reaction;			
KW	multigene family; ligand binding domain.			
OS	Rattus rattus.			
PN	W09211585-A.			
PD	15-OCT-1992.			
PF	06-APR-1992: U02741.			



ID W56671 standard: protein; 428 AA.  
 AC W56671;  
 DT 11-SEP-1998 (first entry)  
 DE Caenorhabditis elegans toc-1 protein.  
 DE toc-1 protein; developmental rate; longevity; cellular physiology;  
 KW cellular metabolism; cancer; clk-1 protein.  
 OS Caenorhabditis elegans.  
 PN W09817823-A1.  
 PD 30-APR-1998.  
 PF 17-OCT-1997: CA07658.  
 PR 18-DEC-1996: US-033195.  
 PR 21-OCT-1996: US-028977.  
 PA (UYMC-) UNIV MCGILL.  
 PI Barnes T, Ewbank J, Hekimi S, Lakowski B;  
 DR WPI: 98-261516/23.  
 PT New Caenorhabditis elegans clk-1 gene - used to obtain human clk-1  
 sequence, useful for, e.g. cancer diagnosis  
 PS Disclosure: page 28; 46pp: English.  
 CC The present sequence represents the Caenorhabditis elegans toc-1  
 CC protein. The toc-1 gene is present within the clk-1 operon of  
 CC C.elegans. The invention claims for the C.elegans clk-1 protein  
 CC (W56670) which is involved in the developmental rate and longevity  
 CC at the cellular physiology level, where clk-1 mutants have a longer  
 CC life and altered cellular metabolism relative to wild-type. The  
 CC clk-1 gene may be cloned to identify related genes, for e.g the human  
 CC clk-1 sequence can be identified and may be useful in the diagnosis  
 CC and/or prognosis of cancer. The invention claims that downregulation  
 CC of expression of clk-1 can be used to increase the life span of animals  
 CC or humans. The invention also claims that if downregulation clk-1  
 CC expression could be targeted to a particular tissue or organ, it could  
 CC lead to a specific physiological slowing down of this tissue/organ and  
 CC a concomitant slower rate of degradation by the ageing process.  
 CC Alternatively, administration of an agent to promote tissue- or  
 CC organ-specific overexpression of clk-1 could allow the physiological  
 CC rates of tissues or organs to be increased, to treat pathological  
 CC conditions causing a slowdown of physiological rate of tissues/organs  
 CC in a patient.  
 SQ Sequence 428 AA;

Query Match 2.7%; Score 103; DB 1; Length 428;

Best Local Similarity 43.9%; Pred. No. 2.95e+01;

Matches 17; Conservative 6; Mismatches 12; Indels 2; Gaps 2;

Db 211 INLCEHFKHDFAMADPYCCLLISVAVFSTWPLSTY 247

OY 197 LRVLCAGIMHNFVAG-VCYLLIS-TWGITMSPLXAY 231

Search completed: Sun Sep 3 12:38:13 2000  
 Job time : 35 secs.

THIS PAGE BLANK (USPTO)





```
SQ Sequence 1147 AA;
Query Match 4.5%; Score 358; DB 1; Length 1147;
Best Local Similarity 25.9%; Pred. No. 6.19e-19;
Matches 195; Conservative 195; Mismatches 19; Indels 72; Gaps 54;

Db 87 LSGQAAAPSLPPOAPTLKXVPSMPATSPGPGIKESVP-LSILQTPQPLPGALL 145
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : :
QY 115 LSG-EGFASMLNPLQSP-PSGGYPADAYRPLSLAQLAAPMTPHOASLSFVNTNGID 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 POSFP-AP-APPQFSSPFLVGLYSPGGFGSTGSPGNTQPLPLASPPGVPPVSLHT 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 QKNFTHAMLSPPHSTMSQPYTEAMCHINGYMSPYDAQGSPGSYQSHQSPPPHH 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 QVGSVVP-QQLLTVTAAPTAAPTVTTVTSIQIQVPLQLQPHFIKADSL-LTAMK-TDGA 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 HHHPMPRIHENPEQVASFSDIADAPETKPTHLVEPQSPKSPQNKKELLRLVNNMSPSEVE 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 TVKA--AGLSPLVSGTIVQTCPLTLYSGGTILATVPLVDAEKLPIINRLAAGSKAPASA 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 RLKNNKSGACSATNGPSRSKEKAIVIQETAEGDED-EDDEDSGETMSQGTIIIVR- 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 QSRGEKRTAHNAIEKRYRSSINDKIELKDLVGVTEAKLNKSAVLRAIDYIRFLQHSNQ 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 351 RPXTERTAHNLIEKRYCSINDRIQLKVLCCGDEAKLSKATLRAIEHIEVEHENQ 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 KLKOENLSLTAVHKSLSKDLVSACSGGNTDVLMEGVKTEVEDTLTPPPSDAGSPFQS 438
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 411 VLKHHVQMRKTLQNNR-L-PY-PE--PIQYTE-Y--SARSPVESPPRNE---R-KR 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 SPLSLGSRGSGSGSDSPDSVFEEDSKAKPEQPSLSHRGMLDRSLALCTLVFLCL 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 SRMS-TT-TPMKNGT-RDGSCKVTLF--AMLLAVLI--FNPIGLLA-G-S-A-IF-- 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 499 SCNPLASLLGARGLPSPDSTTSVYHSPGRNVLGTSRDGPGWAQWLLP-PVWLLNGLLV 557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 503 S-KA-AA--EAP-IASFEGHVRIDDP-D--GTSTRT-LFWECSIINMSYVWFNIMI 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 558 LVSLVLELVYGEVTRPHSGPAYFWRRKQADLDLARGDFQAQAQQLWLALRGLPLP 617
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 553 IYVVKLLIHGDPVODPMSVSWQTFVTTREKARAEKNSGKLDQAQRKFCCLATLDRSLP 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 TSHLDIACSLLWNLIRHLLOBLWGRWLAGRAGLQODCALRVDSASARDAALVYHKLH 677
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 613 SPGVDSVFSVGWECVREHLLWMLGRIYARRRSTTKPVS-V-CRSHAQTAVL-YHEIH 669
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 678 QLHTMG-K-H--T-G-GHLTATNALIALNLAECAG--DAVSATLAETIYVAAALRVK 727
   ||| || : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 QLHLMGITGFEDTIEFSALTGFMSLCVNLAEAGASNDGLPRVYMAQIYISASIQCR 729
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 728 TSLPRAL-HFLTRFLSSAR-QACLAQSGSVPPAMQWLCHPVGRHFFVDGDW-S-VLSTP 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 730 LALPNLLAPFSGYFLRRARRHVRRAPEHVSVHLL-WIEHPATRKFSDAKRLHVLSSK 788
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 784 WESL-Y-SLAGN-PVDPLAQVTLFREHLLER 812
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 789 QKQLRFGSFEDEQLSPARITRTLKVYLLSK 820
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
ID Y07242 standard; protein; 1829 AA.
AC Y07242;
DT 06-JUL-1999 (first entry)
DE Actin-filament binding protein 1-Afadin.
KW Actin-binding protein 1-Afadin; rat; foetal brain tissue; infiltration;
KW metastasis; carcinoma; diagnosis.
OS Rattus sp.
PN EP-905239-A2.
PF 31-MAR-1999.
PR 21-SEP-1998; 307643.
PR 22-SEP-1997; JP-257043.
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
CC (OBAI/) OBAISHI H.
```

```
PA (WADA/) WADA M.
DR WPI; 99-206773/18.
PT New actin filament-binding protein 1-Afadin is useful for diagnosing
PT and treating carcinomas
PS Claim 1; Page 8-18; 27pp; English.
CC This sequence represent an actin-binding protein 1-Afadin derived from
CC rat foetal brain tissue. The isolated protein has a molecular weight
CC of around 205 KD. 1-Afadin and its derivatives may be used to clarify
CC the mechanisms of infiltration and metastasis of carcinoma useful in
CC diagnosing or treating carcinomas.
SQ Sequence 1829 AA;

Query Match 1.9%; Score 151; DB 1; Length 1829;
Best Local Similarity 23.4%; Pred. No. 3.08e-02;
Matches 59; Conservative 68; Mismatches 110; Indels 15; Gaps 14;

Db 1476 QRQETVIRELOPQOQPRTIERRDLOVITTSKEELSGDLSL-PDPWKRDAAREKLEKQOQ 1534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 QHQSPPPHHHHHPMKIHNPE-QVASPSIEDAPETKPTHLVEPOSQPNM-KEEL 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1535 MHI-VDMLSKREIHELQNGDRTAEESDRLRKLMLEWQFQKRLOESKQKDEDDDEEDDV 1593
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 280 LRLVNNMSPSEVERLKNKSGACSATNGPSRSK-EKAAKIVIOETAEGDEDEDSDSG 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1594 DTMLIMORLAEARLARLODEERRRQOOLEMRKR-EVEDRVRQDEEDGRHQEERVKDAE 1652
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 ETMSQGTIIIV-RRPK--TE-RRTAHNLIEKKYRCSINDRIQQLKVLICGDEAKLSKAT 394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1653 EKRRQEGYSRLAEARRRROHEAARLLPEEPGLSRPPLPODY-EPPSOSSAPSPPP 1711
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 LRAIEHIEHVEHENQVLKHHVQMRKTLQ-NNR-LPYPE-PIQYTEYSARSPVESPPSP 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1712 PP-QRNASYLKT 1722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 452 PRNERKPSRMST 463
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
ID W52821 standard; Protein; 597 AA.
AC W52821;
DT 08-SEP-1998 (first entry)
DE Human PRCC-TFE3 construct protein from cell lines UOK120 and UOK146.
KW PRCC; papillary renal cell carcinoma; TFE3; transcription factor;
KW fusion protein; translocation; diagnosis; treatment.
OS Homo sapiens.
OS Synthetic.
FH Key
FT Protein
FT 1. .597
FT /label= PRCC-TFE3
FT /note= "Fusion protein"
FT Misc_difference 1. .156
FT /label= PRCC
FT /note= "papilloma renal cell virus partial sequence"
FT Misc_difference 159. .597
FT /label= TFE3
FT /note= "transcription factor partial sequence"
PN WO9806871-A1.
PD 19-FEB-1998.
PF 13-AUG-1997; G02209.
PR 13-AUG-1996; GB-016986.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI Clark J, Cooper C, Shipley J;
DR WPI; 98-159557/14.
DR N-PSDB; V20957.
PT Diagnosing papillary renal cell carcinoma by detecting gene
PT trans-location - resulting in fusion of TFE3 gene with some other
PT gene, also related vectors, transformed cells, specific binding
PT reagents, peptide(s) encoded by fusions and therapeutic anti-sense
PT sequences
PS Claim 9; Fig 2B; 7lpp; English.
CC This sequence represents a novel fusion protein constructed from a
CC papillary renal cell carcinoma (PRCC) associated protein and the
CC transcription factor TFE3 which is found in cell lines UOK120 and UOK146
```







PR 01-JUN-1994; US-252966.  
 PR 19-SEP-1991; US-756195.  
 PR 23-JUN-1992; US-903710.  
 PR 01-APR-1994; US-222638.  
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 PI Ayer DE, Eisenman RN;  
 DR WPI: 97-258216/23.  
 DR N-PSDB: T70134.  
 PT msin nucleic acids encoding recombinant polypeptide(s) that  
 PT associate with Mad polypeptide - are possible homologues of S.  
 PT cerevisiae general repressor protein  
 PS Example 2; Fig 2D; 11lpp; English.  
 CC This sequence represents Max 14, a basic helix-loop-helix z4pper  
 CC (bHLHZip) protein. Max 11 and Max 14 cDNAs appear to be partial,  
 CC from a Manca (human Burkitt's lymphoma cell line) lambda gt10 library  
 CC overlapping cDNAs. Subsequent isolation of several overlapping cDNAs  
 CC permitted deduction of an apparently complete open reading frame for  
 CC Max that encodes 151 residues. The 9-amino acid insertion found in  
 CC several PCR clones is not indicated in the specification. Max is an  
 CC obligate partner for the DNA binding and transcriptional functions of  
 CC Myc family proteins as well as for the Mad protein. Max is a stable,  
 CC ubiquitously expressed protein which in general does not appear to be  
 CC regulated during mitogenesis, the cell cycle, or differentiation.  
 CC Expression of Mad is closely linked to differentiation in at least two  
 CC distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may  
 CC reflect the repression of transcription of Myc regulated genes by Mad.  
 CC The DNA, vectors and host cells of the invention are useful for the  
 CC recombinant production of msin proteins useful in elucidation of Mad  
 CC repressor functions.  
 SQ Sequence 151 AA;

Query Match 1.6%; Score 125; DB 1; Length 151;  
 Best Local Similarity 24.4%; Pred. No. 2.16e+00;  
 Matches 20; Conservative 29; Mismatches 31; Indels 2; Gaps 2;  
 Db 1 MSDNDIDIEV-SDAKRAHNALEKRRDHDKDSFHSRDSVPSLQGEKASRAQILDKAT 59  
 QY 341 MSQGTITIVRRPKETRTAHNLIEKYRCSINDRIQQLKVLGG-DEAKLSKATLRRAI 399  
 Db 60 EYIQWRRKNHHTQDIDDKR 81  
 QY 400 EHIEVEHENQVLKHHVQEMRK 421

RESULT 11  
 ID W60569 standard; Protein; 346 AA.  
 AC W60569;  
 DT 04-SEP-1998 (first entry)  
 DE Human chromosome 19 derived USF2 gene product sequence.  
 KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;  
 KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;  
 KW multicystic renal dysplasia; renal agenesis; hydronephrosis;  
 KW Von Mayer-Rokitansky-Kuester disorder; bifid ureter.  
 OS Homo sapiens.  
 PN W09815650-A2.  
 PD 16-APR-1998.  
 PF 09-OCT-1997; E05583.  
 PA (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
 PI Frys JPGJ, Groenen PMA, Van De Ven WJM;  
 DR WPI: 98-240833/21.  
 DR N-PSDB: V18130.  
 PT Hydronephrosis gene - useful to treat or diagnose renal diseases and  
 PT disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction  
 PT obstruction, multicystic renal dysplasia or renal agenesis  
 PS Disclosure; Fig 6A-B; 73pp; English.  
 CC This is a USF2 gene product derived from human chromosome 19 USF2 gene. A  
 CC translocation partner to this gene on chromosome 6 is the hydronephrosis  
 CC gene (HNG) product. The HNG gene can be used as a starting point to  
 CC design suitable compounds or techniques for the treatment of renal  
 CC diseases or disorders, or nucleotide probes for diagnosing cells involved  
 CC in renal diseases or disorders. A protein or a fragment encoded by HNG  
 CC gene can be used as a starting point for preparing suitable antibodies

CC for diagnosing cells involved in renal diseases and disorders. The  
 CC products and method can be used to treat or diagnose renal diseases and  
 CC disorders selected from vesical-ureteral reflux, uni or bilateral  
 CC pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal  
 CC agenesis, renal aplasia, hydronephrosis, Von Mayer-Rokitansky-Ruester  
 CC disorder and bifid ureter.  
 SQ Sequence 346 AA;

Query Match 1.5%; Score 121; DB 1; Length 346;  
 Best Local Similarity 34.9%; Pred. No. 4.05e+00;  
 Matches 30; Conservative 19; Mismatches 29; Indels 8; Gaps 6;

Db 229 RTPDRERRAQNHEVERRRRKINNWIQVLSKIPDCNADNSLTGASGGILSKACDYIR 288

QY 350 RPKTERTA-HNLIEKYRCSINDRIQQL-KVLL-CGDEAKL---SKSATLRRRAIEHIE 403

Db 289 ELRQTNQRMQETFEKER-LQMDNEL 313

QY 404 EVEHENQVLKHHVQEMRKTLQ-NNRL 428

RESULT 12  
 ID R77505 standard; Protein; 352 AA.  
 AC R77505;  
 DT 14-APR-1996 (first entry)  
 DE Frog neurogenic differentiation (NeuroD) protein.  
 KW NeuroD; neurogenic differentiation; neuronal growth factor;  
 KW basic helix-loop-helix secondary structure;  
 KW non-neuronal cell differentiation; antigen; drug screening;  
 KW neurodegenerative disease; traumatic injury; gene therapy.  
 OS Xenopus laevis.  
 FH Key Location/Qualifiers  
 FT domain 56..79  
 FT domain /note= "acidic domain"  
 FT domain 118..157  
 FT region /note= "HLH domain"  
 FT region 157..199  
 FT /note= "highly conserved region"  
 PN W09530693-A1.  
 PD 16-NOV-1995.  
 PF 08-MAY-1995; U05741.  
 PR 06-MAY-1994; US-239228.  
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 PA (WEIN/) WEINTRAUB N.  
 PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;  
 DR WPI: 95-404081/51.  
 DR N-PSDB: T05514.

PT Nucleic acid molecule which hybridises with a neuroD HLH domain - is  
 PT used in a method for inducing differentiation of a non-neuronal cell.  
 PS Claim 9; Page 39-40; 50pp; English.  
 CC NeuroD induces differentiation of a non-neuronal cell into a  
 CC neuron. It is a member of a novel protein family and is  
 CC transiently expressed in differentiating neurons during  
 CC embryogenesis. DNA encoding NeuroD may be used in the development  
 CC of probes, in the construction of recombinant cell lines and  
 CC transgenic animals, and in the construction of gene therapy vectors  
 CC for the repair of neuronal defects resulting from traumatic injury  
 CC and neurodegenerative diseases (Alzheimer's disease, Huntington's  
 CC disease, Parkinson's disease). Transformed host cells are used (1)  
 CC as a source of neuronal growth factors, (2) in transient and  
 CC continuous cultures for anti-cancer drug screening, and (3) as  
 CC sources of recombinant NeuroD for use as an antigen in diagnostic  
 CC antibody production.  
 SQ Sequence 352 AA;

Query Match 1.5%; Score 118; DB 1; Length 352;  
 Best Local Similarity 28.3%; Pred. No. 6.47e+00;  
 Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db 33 DLEKKEGELMKEDDEDSLNHHNGENEEDEGDEDEDEDEDDDDDKPKRRGPKKKM 92

QY 285 NMSPSEVERLNKKSGACSATNGPSRKEAKIVIQETAEDEDEDEDESDSGETMSQG 344

```

Db   93 TKARVERFKV-RRKANARENRMHGLNDALDSLRKVVPCYSKYTKLSKETIURLAKNYI 151
      | | | | | | | | | : | | : : | | : : | | | | | | | | | |
Qy   345 TTIIVRPKTERTTAHLNIEKKYRCSINDRIQOL-KVLLC-GDEAKLSKSATLRAIEHI 402

RESULT 13
ID W22437 standard; Protein; 352 AA.
AC W22437;
DT 21-OCT-1997 (first entry)
DE Frog neurogenic differentiation protein (NeuroD1).
KW Neurogenic differentiation protein; NeuroD; neuroD1 gene;
KW transcriptional activator; neuron; pancreas; gastrointestinal;
KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy.
OS Xenopus laevis.
FS Key
FH Location/Qualifiers
FT FT region /label= HLH
FT FT /note= "helix-loop-helix region"
FT FT 157..199
FT FT /note= "highly conserved region characteristic
FT FT of NeuroD proteins"
FT WO9716548-A1.
PN PD 09-MAY-1997.
PF PF 30-OCT-1996; UI7532.
PR PR 02-NOV-1995; US-552142.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA (WEIN/) WEINTAUB N.
PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;
DR N-PSDB: T74888.
PT Nucleic acid encoding neurogenic differentiation polypeptide -
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT development
PS Claim 9; Page 57-58; 8pp; English.
CC Cuman, Mouse and Xenopus neurogenic differentiation proteins
CC (W22436-43) are transcriptional activators involved in neuronal,
CC endocrine and gastrointestinal development. They are active very
CC early in embryo development and may represent an/off switches,
CC possibly reversible, for controlling cell development. They were
CC discovered by expression cloning and screening assays designed to
CC identify basic-helix-loop-helix proteins capable of interacting
CC with the protein product of the Drosophila daughterless gene. Human,
CC mouse and frog neuroD gene sequences (T74887-94) have been
CC identified. Cells expressing NeuroD polypeptides can be used to
CC screen for anticancer agents able to drive terminal differentiation
CC in neural tumours, to screen for (antagonists, or transplanted to a
CC site of neuronal injury or disease, and possibly (since NeuroD can
CC regulate insulin expression) to treat diabetics.
SQ Sequence 352 AA;

Query Match 1.5%; Score 117; DB 1; Length 352;
Best Local Similarity 28.3%; Pred. No. 7.55e+00;
Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db   33 DLEKKEGELMKDEDDSLNHNGHNEENEEDGGDEEDEDDEDDDDQPKRRGPKKXKM 92
      : | | : : | | : : | | : : | | | | | | | | | |
Qy   285 NMSPSEVRLLKNKSGACSATNGPSRSKKAAKVIQTAEAGDEDEDEDDSDGETMSQG 344

Db   93 TKARVERFKV-RMKNANARENRMHGLNDALDSLRKVVPCYSKYTKLSKETIURLAKNYI 151
      | | | | | | | | | : | | : : | | : : | | | | | | | | | |
Qy   345 TTIIVRPKTERTTAHLNIEKKYRCSINDRIQOL-KVLLC-GDEAKLSKSATLRAIEHI 402

RESULT 14
ID W71010 standard; Protein; 352 AA.
AC W71010;
DT 21-OCT-1998 (first entry)
DE Xenopus neuroD1 protein, which is a bHLH protein.
KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW classification; medulloblastoma; frog.
OS Xenopus laevis.
FS Key
FH Location/Qualifiers
FT FT region /label= HLH
FT FT /note= "helix-loop-helix region"
FT FT 157..199
FT FT /note= "highly conserved region characteristic
FT FT of NeuroD proteins"
FT WO9716548-A1.
PN PD 09-MAY-1997.
PF PF 30-OCT-1996; UI7532.
PR PR 02-NOV-1995; US-552142.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA (WEIN/) WEINTAUB N.
PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;
DR N-PSDB: T74888.
PT Nucleic acid encoding neurogenic differentiation polypeptide -
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT development
PS Claim 9; Page 57-58; 8pp; English.
CC Cuman, Mouse and Xenopus neurogenic differentiation proteins
CC (W22436-43) are transcriptional activators involved in neuronal,
CC endocrine and gastrointestinal development. They are active very
CC early in embryo development and may represent an/off switches,
CC possibly reversible, for controlling cell development. They were
CC discovered by expression cloning and screening assays designed to
CC identify basic-helix-loop-helix proteins capable of interacting
CC with the protein product of the Drosophila daughterless gene. Human,
CC mouse and frog neuroD gene sequences (T74887-94) have been
CC identified. Cells expressing NeuroD polypeptides can be used to
CC screen for anticancer agents able to drive terminal differentiation
CC in neural tumours, to screen for (antagonists, or transplanted to a
CC site of neuronal injury or disease, and possibly (since NeuroD can
CC regulate insulin expression) to treat diabetics.
SQ Sequence 352 AA;

Query Match 1.5%; Score 117; DB 1; Length 352;
Best Local Similarity 28.3%; Pred. No. 7.55e+00;
Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db   33 DLEKKEGELMKDEDDSLNHNGHNEENEEDGGDEEDEDDEDDDDQPKRRGPKKXKM 92
      : | | : : | | : : | | : : | | | | | | | | | |
Qy   285 NMSPSEVRLLKNKSGACSATNGPSRSKKAAKVIQTAEAGDEDEDEDDSDGETMSQG 344

Db   93 TKARVERFKV-RMKNANARENRMHGLNDALDSLRKVVPCYSKYTKLSKETIURLAKNYI 151
      | | | | | | | | | : | | : : | | : : | | | | | | | | | |
Qy   345 TTIIVRPKTERTTAHLNIEKKYRCSINDRIQOL-KVLLC-GDEAKLSKSATLRAIEHI 402

RESULT 14
ID W71010 standard; Protein; 352 AA.
AC W71010;
DT 21-OCT-1998 (first entry)
DE Xenopus neuroD1 protein, which is a bHLH protein.
KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW classification; medulloblastoma; frog.
OS Xenopus laevis.
FS Key
FH Location/Qualifiers
FT FT region /label= HLH
FT FT /note= "helix-loop-helix region"
FT FT 157..199
FT FT /note= "highly conserved region characteristic
FT FT of NeuroD proteins"
FT WO9716548-A1.
PN PD 09-MAY-1997.
PF PF 30-OCT-1996; UI7532.
PR PR 02-NOV-1995; US-552142.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA (WEIN/) WEINTAUB N.
PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;
DR N-PSDB: T74888.
PT Nucleic acid encoding neurogenic differentiation polypeptide -
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT development
PS Claim 9; Page 57-58; 8pp; English.
CC Cuman, Mouse and Xenopus neurogenic differentiation proteins
CC (W22436-43) are transcriptional activators involved in neuronal,
CC endocrine and gastrointestinal development. They are active very
CC early in embryo development and may represent an/off switches,
CC possibly reversible, for controlling cell development. They were
CC discovered by expression cloning and screening assays designed to
CC identify basic-helix-loop-helix proteins capable of interacting
CC with the protein product of the Drosophila daughterless gene. Human,
CC mouse and frog neuroD gene sequences (T74887-94) have been
CC identified. Cells expressing NeuroD polypeptides can be used to
CC screen for anticancer agents able to drive terminal differentiation
CC in neural tumours, to screen for (antagonists, or transplanted to a
CC site of neuronal injury or disease, and possibly (since NeuroD can
CC regulate insulin expression) to treat diabetics.
SQ Sequence 352 AA;

Query Match 1.5%; Score 117; DB 1; Length 352;
Best Local Similarity 28.3%; Pred. No. 7.55e+00;
Matches 34; Conservative 22; Mismatches 61; Indels 3; Gaps 3;

Db   33 DLEKKEGELMKDEDDSLNHNGHNEENEEDGGDEEDEDDEDDDDQPKRRGPKKXKM 92
      : | | : : | | : : | | : : | | | | | | | | | |
Qy   285 NMSPSEVRLLKNKSGACSATNGPSRSKKAAKVIQTAEAGDEDEDEDDSDGETMSQG 344

Db   93 TKARVERFKV-RMKNANARENRMHGLNDALDSLRKVVPCYSKYTKLSKETIURLAKNYI 151
      | | | | | | | | | : | | : : | | : : | | | | | | | | | |
Qy   345 TTIIVRPKTERTTAHLNIEKKYRCSINDRIQOL-KVLLC-GDEAKLSKSATLRAIEHI 402

```

```
Best Local Similarity 23.8%; Pred. No. 2.53e+00;
Matches 30; Conservative 25; Mismatches 67; Indels 4; Gaps 4;

Db      42 SHHPQHNNHHHHHHPPAPQPPPPQQOQQPAPQPPOARGAPADDDKGGQPILL 101
       |:| :| | | | | | | | | | | | | | | | | | | | | | | | | |
QY     221 SQHQSPPPHHHHHMPXIHENPEQVSPSIEDAETPKTHLVEPSQSPQNKEELL 280

Db      102 PPSAALDGAKADALGAGPGCGPALAPVPDPEKEKGAGAGEKKGAGEGKGEGGK 161
       |:| :| | | | | | | | | | | | | | | | | | | | | | | | | |
QY     281 RLLVNNSPSEVERLNK - KSGACSATNGP - SRS - KEKAKI VIQET - AEGDEDEDSD 336

Db      162 EGDKN 167
       |:| :| :|
QY     337 SGETMS 342
```

Search completed: Sun Sep 3 12:33:46 2000  
Job time : 48 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sun Sep 3 12:30:56 2000; Maspar time 53.35 Seconds  
Tabular output not generated. 984.052 Million cell updates/sec

Title: >US-09-332-522B-2  
Description: (1-1113) from US09332522B.pep  
Perfect Score: 8012  
Sequence: 1 MNEFEFGDVPMSDPFLSLVT.....AQPDAHLHTLVKLTSMOL 1113

Scoring table:  
PAM 150  
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pirs4  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 55.341; Variance 113.686; scale 0.487  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description	Pred. No.
1	443	5.5	1139	2	B54962	sterol regulatory ele 3.72e-55
2	421	5.3	1141	2	A54962	sterol regulatory ele 4.47e-51
3	387	4.8	1133	2	A54164	sterol regulatory ele 7.85e-45
4	375	4.7	927	2	A48085	transcription factor 1.20e-42
5	322	4.0	1113	2	T13348	transcription factor 3.80e-33
6	265	3.3	391	2	PD0035	sterol regulatory ele 2.84e-23
7	149	1.9	419	2	A40728	microphthalmia-assoc 3.81e-05
8	149	1.9	419	2	I38024	MTF protein - human 3.81e-05
9	149	1.9	520	2	T14752	microphthalmia-assoc 3.81e-05
10	141	1.8	320	2	B36289	CemYOD protein - Caen 4.61e-04
11	143	1.8	446	2	A47029	transcription factor 2.49e-04
12	145	1.8	476	2	A54724	transcription factor 1.34e-04
13	143	1.8	536	2	A34596	transcription factor 2.49e-04
14	133	1.7	309	2	S08343	nodulin precursor - s 5.16e-03
15	137	1.7	324	2	A36289	CemYOD protein 1 - Ca 6.27e-04
16	140	1.7	413	2	S28066	esc1 protein - fission 5.16e-03
17	133	1.7	473	2	C64739	YadO protein - Escher 5.16e-03
18	125	1.6	151	2	B38431	Myc-binding factor Ma 5.27e-02
19	128	1.6	260	2	A41123	myogenic transcriptio 2.23e-02
20	126	1.6	310	2	S15325	transcription factor 3.96e-02
21	132	1.6	317	2	A48080	basic helix-loop-heli 6.93e-03
22	127	1.6	338	2	I56893	transcription factor 2.98e-02
23	131	1.6	387	2	A47446	HNF-3/fork head fam1 9.30e-03

24	130	1.6	391	2	T04477	b12-1 protein - barle 1.25e-02
25	131	1.6	405	2	J02147	ORH1 protein - maize 9.30e-03
26	131	1.6	410	2	J02148	ORH2 protein - maize 9.30e-03
27	127	1.6	417	2	T15174	hypothetical protein 2.98e-02
28	132	1.6	434	2	S62168	hypothetical protein 6.93e-03
29	132	1.6	451	2	A55909	transforming protein 6.93e-03
30	127	1.6	514	2	A35658	transcription factor 2.98e-02
31	126	1.6	675	2	T03744	MYOD protein inhibito 3.96e-02
32	130	1.6	1110	2	A43253	large tr-1 protein - 1.25e-02
33	131	1.6	1212	2	T13804	shs protein - fruit f 9.30e-03
34	125	1.6	1329	2	A64828	cell division protein 5.27e-02
35	125	1.6	2248	2	D42088	adenylyate cyclase (EC 5.27e-02
36	123	1.5	212	2	S42396	USF2 protein - mouse 9.29e-02
37	123	1.5	234	2	I54074	USF2 protein - mouse 9.29e-02
38	123	1.5	346	2	A55111	transcription factor 9.29e-02
39	123	1.5	346	2	A56522	upstream stimulatory 9.29e-02
40	124	1.5	405	2	J05175	seed storage protein 7.01e-02
41	124	1.5	480	2	JH0672	brain factor 1 protei 7.01e-02
42	123	1.5	552	2	H69126	chaperonin - Methanob 9.29e-02
43	124	1.5	790	2	T01537	S-receptor kinase (EC 7.01e-02
44	122	1.5	974	2	T04910	hypothetical protein 1.23e-01
45	123	1.5	3164	1	WPBBH6	UL36 protein - human 9.29e-02

ALIGNMENTS

RESULT 1  
ENTRY B54962 #type complete  
TITLE sterol regulatory element binding protein 2 precursor - Chinese hamster  
ORGANISM #formal name Cricetus griseus #common name Chinese hamster  
DATE 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 04-Sep-1998

ACCESSIONS B54962  
REFERENCE A54962  
#authors Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.  
#journal Genes Dev. (1994) 8:1910-1919  
#title Sterol-resistant transcription in CHO cells caused by gene rearrangement that truncates SRBP-2.  
#cross-references M01D:95047343

#accession B54962  
#status preliminary  
#molecule\_type mRNA  
##residues 1-1139 ##label YAN  
#cross-references GB:U12330  
#note 493 Ser was also found  
CLASSIFICATION #superfamily sterol regulatory element binding protein  
KEYWORDS DNA binding; membrane protein  
SUMMARY #length 1139 #molecular-weight 123681 #checksum 3153

Query Match 5.5%; Score 443; DB 2; Length 1139;  
Best local similarity 25.5%; Pred. No. 3.72e-55;  
Matches 193; Conservative 208; Mismatches 280; Indels 75; Gaps 56;  
Db 324 PREGRRRTNHNIEKRYSSINDKIEIKDVMGDAMHRSYRKADYIKYIQYVNH 383  
332 PK-TERRAHLNDIEKRYKCSINDRTOQLKVLGDEALSKSATLRRAIEH---EVEH 407  
384 KLROENMVLKLANOKNKLKIDIGSLVSDVDLIKIEDFNQNVILMSPASDSGQAQFS 443  
408 ---ENQVLK--HHVEQKRKTLQNNRLPYPE-PIQYTVYSARSPESSPPRNRRK--R 458  
Db 444 PYSIDSEPGSPLLDDAKYKDEPPVALGAVDRSRILCVTLFGSLFNPJLSILQMG 503  
459 S-RM-S-TTTPM-KNGT-RD--GSSKVTL-F---A-MLAVLINFPIGLILGSAIFSKAA 506  
Db 504 AHDTHQHPYSGSGRSVLSGSG-GWF-D-WMMPTLLMLVNGVIVSVFKLVHGEP 560  
507 AEAPLASPEF-HGRVIDPDGSTRITLFWESGIIMMSYVWFNTIMITYVKKLIHODP 565  
Db 561 VIRPHTRSVTFWHRKROADDLARGDFMAAANLQTCISVGRALPTSRDLACLSWN 620  
566 VQDFNSVSWQTFVTTRREARALNSGNLKDQGRFCBELATLDSLSPGVDVSFVQWE 625

Query Match:	5.3%:	Score 421:	DB 2:	length 1141:
Best Local Similarity	25.1%	Pred. No. 4,47e-51:		
Matches 190:	Conservative 210:	Mismatches 280:	Indels 77:	Gaps 59:
Db	326	PRGEGRTHNTEKRRSSINDKIIELKDLVWGTDAAKHKSGVLKAIIDYIKYLOOVNH	385	
Qy	352	PK-TERRTAHNTEKRRKCSINDRIQDLVLLCGDEAKLSKATLRRRAIEHI--DEVEH	407	
Db	386	KLRQEMVYLKANOKNKKLIGDLGSLVDNEVDKIEDENOVNVLMSPPASDSGSGAGFS	445	
Qy	408	----ENQYIK--HNYEQMKTIQNNRPLRPE-PIQYTERSAKSPVSSSPSPRNEKK--R	458	
Db	446	PYSIDSEPGSLDDAKYKDEPDSPVALGMYDRSHILLVTLFCLSFENPLTSLLOWGG	505	
Qy	459	S- <u>RM-S-TTTPM-KNGT-RD--GSSKVTL-F--A-MLTAVLIFNPIGLAGSAIFSKAA</u>	506	
Db	506	AHDSOHPHSGSGRSLVSPSSSG--GWF-D-WMMPTLLMLVNGVVLSEVYLLVHGEF	562	
Qy	507	AAPAPLAPPE-HGRVIYDDPGTSTRILFMEGSIIINNSYVWVNIILIIYVYKLLHGD	565	
Db	563	VIRPHSRSSVTFWRHRRQADLDLADGFAAAAANIOTCLAVLGRALPTSRDLACLSMN	622	
Qy	566	VODFMSVSMQGTVTTREREKARALNSGLMDAQRKFECECATIDRSPLSGVSVFSGVGE	625	
Db	623	VIRYSLOKRLVRLWLKAYFOCRATPATGEDEDAKTSARDPAALAVYRLHOLH--IT	679	
Qy	626	CYRH-LTN-WL-WI-GR-YIARRRSTKPV-SVACRSHAQAVL-YHEIHQLHLMGIT	677	
Db	680	GKLP-A-GSAC-SDVHMLCAVNLAECA--BEKIPSPSLVEHLHTAAMGLKTRCGGKLG	733	

[illegible]

#cross-references	MUID:94274723
#accession	A54164
#status	preliminary
#molecule_type	mRNA
#residues	1-1133 ##label SAT
#cross-references	GB:U09103; NID:9516002; PIDN:AAA20085.1; PID:9516003
CLASSIFICATION	#superfamily sterol regulatory element binding protein
SUMMARY	#length 1133 #molecular-weight 120464 #checksum 4386
Query Match	4.8%; Score 387; DB 2; Length 1133:
Best Local Similarity	29.4%; Pred. No. 7.85e-45;
Matches	149; Conservative 132; Mismatches 171; Indels 54; Gaps 41;
Db	313 OSRGKRTAHNAIEKRRYSINDKIVELKDLVYTEKLIKSAVLKRAIDYIFELQHSNO 372
Oy	331 RPKETRRRAHNLIEKKYKCSINDRIQOKVLLGDEKELKLSATLRRAIEHIEVEHENQ 410
Db	373 KLGQENLIRNAHRSKSLKDLVACAGSAGGTVDVAMGVKPEVVDLTTPPSDAGSPSOS 432
Oy	411 VLKHVEDEMRKTLQNNR-L-PY-PE--PIQYTELSARS--P-VES-SPSP-P-RNEKRS 459
Db	433 SPLSLGSGSSSGSDSEPPDSPEVEDSOVKAKQRLHSHQMDRSLALCALVLCITCNPL 492
Oy	460 R-MSTTTP-MKNGTRDGSKYTLER-MLAVLINPPIGLLAGSAI-F-SKA-A-AE-APi 511
Db	493 ASLQGWGIPGRSSASGAHSSGRSMLAESDGSNMWOMLPLPWLANGLVACIALL 552
Oy	512 ASPPEHG-RVIDDPDGT--ST-RILF-----WEGSI-IN--WS-YVWFNLMTIYYVKL 559
Db	553 FVYGEPTVRPTSPAVH-FWRHRKQADLDLRGDPFAQAQOLWIALQALGRPLPTSLMD 611
Oy	560 LIHGDPV-QDEMVSWSQFVTFREKARLNLSGULKDAQRKFCCLATLDRSLPSPEVDS 618
Db	612 ACSLLMLIRLRLQRLWYGRMLARAGGLRDRDCLRMDARASADALVYHKLQHLHMG 671
Oy	619 VESVMECEVRLRLWMLWGRYIARRRRRTTSPVS-V-CSSHQTAVL-YHEIHQHLMG 675
Db	672 --KY-T-G-GHILASNLALNALNECAG--DASVMTLAEIYYAALRYKTSIPRA 721
Oy	676 ITGNEDITEPSALTGLFMSLCAYNLALAEAGASDGLPRAYMAQIITYSASIQCLALPNL 735
Db	722 L-HLLREFLLSAR-QACLAOSGVPLAMQWLCHPVGRREFVDDMAVHG-AP-OESL-Y 776
Oy	736 LAFPFSGYFLRRARRHVRAREHSVS-HLWIFPRATRKFSMDAKRLRHVLLSSQKQLRF 794
Db	777 -S-VAGNEVDPDLAQVTRLFCHELLER 800
Oy	795 GSFEDEQLSPLARIRITLTKVLLSK 820
RESULT	4
ENTRY	A48085 #type complete
TITLE	transcription factor ADD1 - rat
ORGANISM	formal_name Rattus norvegicus #common_name Norway rat
DATE	26-May-1994 #sequence_revision 26-May-1994 #text_change
ACCESSIONS	A48085
REFERENCE	A48085
#authors	Tononcelz, P.; Kim, J.B.; Graves, R.A.; Spiegelman, B.M.
#journal	Mol. Cell. Biol. (1993) 13:4753-4759
#title	ADD1: a novel helix-loop-helix transcription factor associated with adipocyte determination and differentiation.
#cross-references	MUID:93330269
#accession	A48085
#status	preliminary
#molecule_type	mRNA
#residues	1-927 ##label TON
#cross-references	GB:U16995
CLASSIFICATION	#superfamily sterol regulatory element binding protein
KEYWORDS	leucine zipper
SUMMARY	#length 927 #molecular-weight 98517 #checksum 1870

Query Match	4.7%	Score 375	DB 2	Length 927
Best Local Similarity	30.2%	Pred. No. 1,20e-42		
Matches	155	Conservative 125	Mismatches 180	Indels 53
Gaps	37			
Db	282	GALSSAOSGKGRANHAIEKRRSSINXIVELKDLVGVTEAKLKSALVRKAIDYR	341	
QY	344	GTITIVRRPKTERTAHLIEKTKRCSINDIIOQLKVLCCGDEKLSKSAITLRRAIEHT	403	
Db	342	FLIHSNOKLKKDENTLRISAKHSKSLKLDIVSAGSGGSDIVSMEGMKPEVEVITLPPESDA	401	
QY	404	EVEHENQVLKHHVEMGRKTLQNN--R-LPY-P-EDIQTEVS--ARSP--VES--SPSP-R	453	
Db	402	GSPSSSPLSLGASRSGSSGGSDSEPDPAEPDNOYKAKQLRPLSHSRGMLDPSRLALCVLF	461	
QY	454	NEKRKR--MSTTTP--MKNGTRDGSCKYTLFA--MLL-AV-LIFNFIIGLGS--A--IFS--	503	
Db	462	LCITGNPLASLFGWMLPPDASGVHSSGSMLEASRPGSMNTQVLLPLVLAAGL	521	
QY	504	KAAAPAPLAPSPHGRV--DDPDGT--ST-RTLF---WEST--IN--MS-YWVENILM	551	
Db	522	VLACALLFVYGEPTRPHSGPAVHFMRKROADLDLARGDFAQAQOLMALQALRPL	581	
QY	552	IIVVVVKKILHGDPPQDPMYSWQTFYTTREKARAEINSGLKDAQKFCFCALATDRSL	611	
Db	582	PLSNLDLACSLIMLVRLQLRWGKRLAQAGGLORDYRLRKDAPASAPDAVYHKL	641	
QY	612	PSPGDVSEVSGWECVRLMLMIGRIARRRSTK--PVSVCVSHAO--TAVLYHEI	668	
Db	642	HOLHMG--KYTGGH-LVA-SN--ALSALNLECG--DAISMRTLEIYVAALRV	691	
QY	668	HOLHMGITGNFEDTIEPSALTGLFMSICAVNLAEAGASNDGILPRAVMQIYISATQC	728	
Db	692	KTSLPRAL--HFLTRFEFLSSAR-QACLAQSGAPLAWOMLCHPGHREFVGDMAVHG--AP	748	
QY	728	RALPNLLAPFSGYFLRARARHVRAPRHSVS--HLMIHPAPARKMSAKRLREHLSS	787	
Db	749	-QESL-Y-S-VAGNPVPLAQVTLFCEHLIER	777	
QY	788	KOKQLRFGSFEDEQLSLARIRITLKYLLSK	820	
RESULT	5			
ENTRY	T13348	#type complete		
TITLE	transcription factor HLM106 - fruit fly (Drosophila			
ORGANISM	melanogaster)			
DATE	13-Aug-1999	#formal name Drosophila melanogaster		
ACCESSIONS	20-Sep-1999	#sequence_revision 13-Aug-1999		
REFERENCE	T13348			
#authors	Theopold, U.; Ekengren, S.; Hultmark, D.			
#journal	Proc. Natl. Acad. Sci. U.S.A. (1996) 93:1195-1199			
#title	HLM106, a Drosophila transcription factor with similarity to			
	the Drosophila steric receptor-like element binding protein.			
	Accession			
	T13348			
	accession			
	T13348			
	status			
	##molecule_type mRNA			
	##residues 1-1113			
	##cross-references EMBL:U38238; NID:g1079655; PTD:g1079656;			
	PIDN:AAA97864.1			
CLASSIFICATION	#superfamily steryl regulatory element binding protein			
SUMMARY	length 1113 #molecular-weight 124513 #checksum 539			
Query Match	4.0%	Score 322	DB 2	Length 1113
Best Local Similarity	24.6%	Pred. No. 3.80e-33		
Matches	192	Conservative 204	Mismatches 305	Indels 81
Gaps	68			
Db	261	SPAAEVOGKYPINVOGKYEKVRSAHNAIERFYSINDIKELKNLVGEOAKLKS	320	
QY	337	SGEYMS-QGTTIIVR-PRKT-E-RTAHNLIEKTKRCSINDRIQQLKVLCCGDEAKLSKS	392	
Db	321	AVLRKSIDKIRDLQRHDLKAEQLQRLQRELMAWDGSKVYDLQGLTRPGRAKKRRSS	380	

QY	393	ATIRAIHIEHVEHENQVTK--HHVE-Q-M-RK-T-LQNN-RLPY-P--EPIDYTESS	440
Db	381	QTFTTDACTIPRDESDPSLSPHMSDISLPPSPYGGSTACSSSSSSSSNEEPLVPPSM	440
QY	441	ARSPRESSPSPR-NERKRS--RM-STTT-PMKN-GTRDGS-SKTYLPAMLAVLIFNFI	493
Db	441	RGMAHSHRLGLCMFMFALLAVNPFETFLQRGYDSNDLDGMSGGORILISYVDEEGFAY	500
QY	494	-GLLA-GS-AL--FEKKA-AEAPLISPEHGHVYIDPD-GT-ST--RTL-FW-EG-SI-I	539
Db	501	WOOSWIMLNFNTLMAGCLVKLIVGDP-Q-LDAQT-DAYCOHROADFEFSOGSSQAY	557
QY	540	-NMS-YVWVFNTIMIIYVVVVKLLIHGDPPQDEMSYSWQTFVTTREKARALNSGMLKDAQ	597
Db	558	AGYLNCIMFGISLPASRLECTLOTTWQFLRFLRLMIGRLVSRSGGLFSNMASRKOA	617
QY	598	RKFCCLATLDRSLSPSGDVSVFSGWECVRLHMLMWLGRIARRRRTTKPVSVGRS	657
Db	618	LASARELALFRLNLOLTGN--GS--RGDNGIMMALFASNMAEVA--HNLTLPRETI	671
QY	658	HAQTAVLYHEIHQHLMTGTTGNEFEDTIEPSALTGLFMSLCANVLAEMAGASMDGILPRAV	717
Db	672	C-IHVTALRMKRSAPKWLQDFPARYMSRAHOEGGRTTATBOT-OELMAFTAYGYR-	728
QY	718	AOIYISASIOCLRALPNLLAPFSGYFLRRARRH-R-RAPEHSVHLLIMPHPATRKPM	775
Db	729	CATHTFTYDLSBSSGDDGFTFLRN-PCDPAHVIKQYREHLLFSSIOCLVAGAKHSGSL	787
QY	776	SDAKRLEHVLSSKQKOLRFGSFVEDEQSLPIARITTKVLLSKLYOELVA-GDEIFTK	834
Db	788	PTSSVSGAEQLOQO-Q--HSGTIVSNV-LKTYSL-LKDT-LMAD-ED--ERDTNV--VM	836
QY	835	NVERLINDNRDDEVDVVDVSRLLVTISOCALITNKDSARFGTPIISNGDACCM	894
Db	837	WADVLETAHVHMLGEDTLAEOULYGRIKQPTLOQCGENDHLPKALHAYLRAMKMTILKN	896
QY	895	WTHVLTGCIYGRSNKNELARQHSYLSIRNCPKIL--TDMLGLAVG-HA-LCARAKICIDR	950
Db	897	GNALDKSLKOLVNIICDESSVLEOECTVYNRITDA-KGIKLLEFOLLTCOMLLEFTRALME	955
QY	951	-DS-PK-VSOYVCHITKKSLESRLRFTSSRSAGVSGIQESTRRMAEWMINSLLDADR	1007
Db	956	LEHNMEDGFE-QY-PGE-VLEK-FOTDINLARNIVENIPASORITLYEAVCRILMGA	1011
QY	1008	-SNL-FASKPYWTQGFKGSTSTLYQEAANYNY-ALINSTRDCMRLLFYELTCMLNGA	1064
Db	1012	SP 1013	
QY	1065	NP 1066	
RESULT	6		
ENTRY	PD0035	#type complete	
TITLE		sterol regulatory element-binding protein-1 isoform - mouse	
ORGANISM		#formal_name Mus musculus #common_name house mouse	
DATE		23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Aug-1999	
ACCESSIONS	PD0035		
REFERENCE	PD0035		
#authors		Inoue, J.; Sato, R.	
#journal		Biosci. Biotechnol. Biochem. (1999) 63:243-245	
#title		A novel splicing isoform of mouse sterol regulatory element-binding protein-1 (SRBP-1).	
#cross-references		MUID:99161303	
#accession		PD0035	
#status		preliminary	
#molecule_type		mRNA	
#residues		1-391 #label INO	
#cross-references		DDbj:AB017337; NID:g4240011	
KEYWORDS		transcription	
SUMMARY		length 391 #molecular_weight 39912 #checksum 4894	

	Query Match	3.3%	Score 265;	DB 2;	length 391;
	Best Local Similarity	50.0%	Pred. No.	2.84e-23;	
Matches	39;	Conservative	18;	Mismatches 21;	Indels 0; Gaps 0;
Dd	268 OSRGKPTAHNAIEKKRPRSSINDKIIVELKDVIQVTEAKLNKSAYLRKAIDYIRFLCHSNQ	327 :::            :                 ::   :::            :                 ::	Oy	351 RPKETRRRAHLLEIKRTGCSINDNIQLTKVLCDDEAKLSATLRAAIEHVEEHENQ	410
Dd	328 KLQEENTLTLSAAHSKSL	345   : :   :   :	Oy	411 VLKHNVDEMKTLLNNRL	428

```

RESULT      7
ENTRY       A40728      #type complete
TITLE       microphthalmia-associated transcription factor mi - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        19-May-1994 #sequence_revision 19-May-1994 #text_change
10-Sep-1997

ACCESSIONS  A40728: 149245; S36645
REFERENCE    A40728
#authors     Hodgkinson, C.A.; Moore, K.J.; Nakayama, A.; Stellingrimsen,
#reference    E.; Copeland, N.G.; Jenkins, N.A.; Arnheiter, H.
#journal      Cell (1993) 74:335-404
#title        Mutations at the mouse microphthalmia locus are associated
              with defects in a gene encoding a novel
              basic-helix-loop-helix-zipper protein.
#cross-references MUID:93345026
#accession   A40728
#status       preliminary
#molecule_type mRNA
#residues     1-419 #label HOD
#cross-references GB:223066; NID:9396739; PID:9396740
REFERENCE    A48021
#authors      Hughes, M.J.; Lingrel, J.B.; Krakowsky, J.M.; Anderson, K.P.
#journal      J. Biol. Chem. (1993) 268:20687-20690
#title        A helix-loop-helix transcription factor-like gene is located
              at the ml locus.
#cross-references MUID:94012591
#accession    A48021
#status        preliminary
#molecule_type DNA
#residues      238-286 #label HUG
#cross-references GB:L22958
REFERENCE    I49244
#authors      Stellingrimsen, E.; Moore, K.J.; Lamoreux, M.L.;
              Ferre-D'Amare, A.R.; Burley, S.K.; Sanders Zimring, D.C.;
              Snow, L.C.; Hodgkinson, C.A.; Arnheiter, H.; Copeland,
              N.G.; Jenkins, N.A.
#journal      Nature Genet. (1994) 8:256-263
#title        Molecular basis of mouse microphthalmia (ml) mutations helps
              explain their developmental and phenotypic consequences.
#cross-references MUID:95179171
#accession    I49245
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues      1-16 #label RPS
#cross-references EMBL:019875; NID:9642572; PID:9642573
GENETICS
#gene          ml
KEYWORDS      DNA binding; transcription factor
SUMMARY       #length 419 #molecular_weight 46768 #checksum 1668

Query Match      1.9%  Score 149;  DB 2;  Length 419;
Best Local Similarity 31.2%  Pred. No. 3,81e-05;
Matches 24;  Conservative 27;  Mismatches 21;  Indels 5;  Gaps 4.

DB      204 OKKDHNLEERRRPNIDRIKGTLPKSNDDPMKNGKTIKASVDYIRKLQRFQOR 263
       355 ERRAAHNLEKKYKCSINDRIQQLKVLIC-GDEAKL--SRSATIRRAIEHIEVEHENOV 411
DB      264 AKD-LENNRQKLEHANR 279

```



OY 412 LKHVHQMKRTLQN-NR 427

RESULT 8

ENTRY 138024 #type complete

TITLE MITF protein human

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 29-Aug-1997

ACCESSIONS 138024

REFERENCE 138024

#authors Tachibana, M.; Perez-Jurado, L.A.; Nakayama, A.; Hodgkinson, C.A.; Li, X.; Schneider, M.; Mikl, T.; Fex, J.; Francke, U.; Arnheiter, H.

#journal Hum. Mol. Genet. (1994) 3:553-557

#title Cloning of MITF, the human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1-p12.3

#cross-references MUID:94348499

#accession 138024

#status preliminary; translated from GB/EMBL/DBJ

#molecule\_type mRNA

#residues 1-419 ##label RES

GENETICS

#cross-references EMBL:229578; NID:9468496; PID:9468497

#gene GDB:MITF

#cross-references GDB:214776; OMIM:156845; OMIM:193510

SUMMARY #map\_position 3p14.1-3p12 #length 419 #molecular\_weight 46938 #checksum 307

Query Match 1.9%; Score 149; DB 2; Length 419; Best Local Similarity 31.2%; Pred. No. 3.81e-05; Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 204 OKDNHNIERRRRNINDRIKELGTLIPKSNDDPMKMKGTILKASVDYIKLQREQOR 263

OY 355 ERRAHNIIEKKYKCSINDRIQQLVLLC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411

DB 264 AKE-LENROKLEHANNR 279

OY 412 LKHVHQMKRTLQN-NR 427

RESULT 9

ENTRY 114752 #type complete

TITLE microphthalmia-associated transcription factor, MITF-A - human

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Sep-1999

ACCESSIONS 114752; JEO207

REFERENCE 218180

#authors Koehner, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

#submission submitted to the Protein Sequence Database, August 1999

#accession 114752

#status preliminary

#molecule\_type mRNA

#residues 1-520 ##label KOE

#cross-references EMBL:AL110195

#experimental\_source adult uterus; clone DKFZp568B217

REFERENCE JEO207

#authors Amai, S.; Fuse, N.; Yasumoto, K.; Sato, S.; Yajima, I.; Yamamoto, H.; Udono, T.K.; Durlu, Y.; Tamai, M.; Takahashi, K.; Shibahara, S.

#journal Biochem. Biophys. Res. Commun. (1998) 247:710-715

#title Identification of a novel isoform of microphthalmia-associated transcription factor that is enriched in retinal pigment epithelium.

#cross-references MUID:98321192

#accession JEO207

#molecule\_type mRNA

#residues 1-520 ##label AMA

#cross-references DBJ:AB006909

COMMENT This protein plays a important role in differentiation of retinal pigment epithelium and regulates melanogenesis in retinal pigment epithelium.

GENETICS

#note DKFZ568B217.1

SUMMARY #length 520 #molecular\_weight 58162 #checksum 6813

Query Match 1.9%; Score 149; DB 2; Length 520; Best Local Similarity 31.2%; Pred. No. 3.81e-05; Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 305 OKDNHNIERRRRNINDRIKELGTLIPKSNDDPMKMKGTILKASVDYIKLQREQOR 364

OY 355 ERRAHNIIEKKYKCSINDRIQQLVLLC-GDEAKL--SKSATLRRAIEHIEVEHENOV 411

DB 365 AKE-LENROKLEHANNR 380

OY 412 LKHVHQMKRTLQN-NR 427

RESULT 10

ENTRY B36289 #type complete

TITLE CeyYOD protein - Caenorhabditis elegans

ORGANISM #formal\_name Caenorhabditis elegans

DATE 08-Mar-1991 #sequence\_revision 08-Mar-1991 #text\_change 23-Feb-1997

ACCESSIONS B36289

REFERENCE B36289

#authors Krause, M.; Fire, A.; Harrison, S.W.; Priess, J.; Weintraub, H.

#journal Cell (1990) 63:907-919

#title CeyYOD accumulation defines the body wall muscle cell fate during Caenorhabditis elegans embryogenesis.

#cross-references MUID:91077929

#accession B36289

#status preliminary

#molecule\_type DNA

#residues 1-320 ##label KRA

#cross-references GB:M37497

#note The authors translated the codon TTG for residue 258 as Lys

KEYWORDS DNA binding; nucleus; transcription regulation

SUMMARY #length 320 #molecular\_weight 36032 #checksum 2241

Query Match 1.8%; Score 141; DB 2; Length 320; Best Local Similarity 25.6%; Pred. No. 4.61e-04; Matches 32; Conservative 35; Mismatches 49; Indels 9; Gaps 7;

DB 110 EKSTPNATELIQSRYVDSOHEDTTSTAGAGVG--PRRTKIDRRKAATMRERRLR-KV 167

OY 313 EKAAKIVIOETAEDEDEDESDSGETMSQSTIIIVRPKFKRTAHLIKK-YRCGI 371

DB 168 NEAFEVKQRTCPNPORLPKVELLSAIDYINLNR--ML-QQAGKTKIMEQNHLO 223

OY 372 NDRIQQLKVLGCD-EAKLSKATLRRAIEHIEVEHENOV LKHVHQMKRTL-QNNRLP 429

DB 224 MTQOI 228

OY 430 YPEPI 434

RESULT 11

ENTRY A42029 #type complete

TITLE transcription factor E3 - mouse

ORGANISM #formal\_name Mus musculus #common\_name house mouse

DATE 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 29-Jan-1999

ACCESSIONS A42029

REFERENCE A42029

#authors Roman, C.; Matera, A.G.; Cooper, C.; Artandi, S.; Blain, S.; Ward, D.C.; Calame, K.

#journal Mol. Cell. Biol. (1992) 12:817-827

```

#title      mTFE3, an X-linked transcriptional activator containing basic helix-loop-helix and zipper domains; utilizes the zipper to stabilize both DNA binding and multimerization.
#cross-references MUID:92123207
#accession   A42029
##status     preliminary
##molecule_type mRNA
##residues    1-446 ##label ROM
##cross-references GB:S76673; NID:g243439; PID:g243440
##note        sequence extracted from NCBI backbone (NCBIN:76673, NCBIP:76674)
SUMMARY      #length 446 #molecular-weight 47891 #checksum 8873

Query Match          1.8%; Score 143; DB 2; Length 446;
Best Local Similarity 31.2%; Pred. No. 2,49e+04;
Matches 24; Conservative 26; Mismatches 22; Indels 5; Gaps 4;

Db      219 OKKDNIILERRRRRINRIKELGFLIPKSNDPEMKRNKGTLIKASVDYIRKLQKEQR 278
       +:: |||||: | :|||: | : : :::: |:|: | :|: | :|: | :|: | :|: |
Oy      355 ERTAHNLLEKKTKRCINDRIQLKVLC-GDEAKL--SKSALIRRAIEHIEVEHENOV 411
               -:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|

Db      279 SKD-LESRQRSLSQANR 294
           | :-::-|--| 
Oy      412 LKHVHQMRKLT-QNNR 427

RESULT      12
ENTRY TITLE      A54743 #type complete
ORGANISM         transcripition factor HFKI - human
DATE             #formal_name Homo sapiens #common_name man
                18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
                12-Sep-1997
ACCESSIONS       A54743
REFERENCE         AS4743
AUTHORS           Murphy, D.B.; Wiese, S.; Burfield, P.; Schmundt, D.; Mattei,
                  M.G.; Schulz-Schaeffer, W.; Thies, U.
                  Genomics (1994) 21:551-557
                  Human brain factor 1, a new member of the fork head gene
                  family.
                  #citations MUID:95048332
                  #accession A54743
                  #status    preliminary; nucleic acid sequence not shown
                  ##molecule_type RNA
                  ##residues    1-476 ##label MOR
                  ##cross-references GB:X74142
GENETICS          GDB:FKNL4; HBFI:1; HFK1
#gene            ##cross-references GDB:433550
CLASSIFICATION    map_position 1q412-1q412
                  #superfamily unassigned fork head proteins; fork head
                  DNA-binding domain homology
FEATURE          169-260 #domain fork head DNA-binding domain homology #label FHD
SUMMARY          #length 476 #molecular-weight 50830 #checksum 5990

Query Match          1.8%; Score 145; DB 2; Length 476;
Best Local Similarity 29.1%; Pred. No. 1.34e+04;
Matches 37; Conservative 27; Mismatches 55; Indels 8; Gaps 8;

Db      39 HNNSHNPNONNHHNNHHNRPRARPORRRARAOORRRPRLAROGAAOSNDEKRPQ 98
       |||::|:|||::| | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Oy      223 HNOSEPPHHNNHHNMKRIGENHPROVAVSLEDARETPTLL-VEROSPSPQNKKDELIR 281
               -:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|

Db      99 LHL-LPTTNHRPSAGKAAGCCSRGELGVGRDEKKGAGAAGEEGKGAEGSGDEGG 157
       +::|::|::| | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Oy      282 LLVNMSFSEVERL-KKKKSACS-ATN-GP-SRS-KRKAKIVIQEI-AEDDEDDEDSDS 335
               -:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|-:-|

Db      158 KGECKN 164
           || ::
Oy      336 DSGETMS 342

RESULT      13
```

```

ENTRY          A34596      #type fragment
TITLE          transcritpion factor E3 - human (fragment)
ORGANISM       Homo sapiens #common_name man
DATE           06-Jul-1990 #sequence_revision 09-Oct-1992 #text_change
                                     24-Sep-1998
ACCESSIONS     A34596; S10379
REFERENCE       A34596
#authors       Beckmann, H.; Su, L.K.; Kadesch, T.
#journal       Genes Dev. (1990) 4:167-179
#title         TF3: a helix-loop-helix protein that activates transcription
                through the immunoglobulin enhancer muB3 motif.
#cross-references MUID:90249724
#accession     A34596
##status       preliminary
##molecule_type mRNA
##residues     1-536 ##label BEC
##cross-references EMBL:X15130; NID:g37061; PID:e20907; PID:g135343
#note          the authors translated the codon ACC for residue 433 as
                Ser, and GAG for residue 472 as Gln

GENETICS
#gene          GDB:TFE3
#cross-references GDB:125670; OMIM:314310
#map_position  Xp11.23-Xp11.22
KEYWORDS       DNA binding; transcription factor
SUMMARY        #length 536 #checksum 8782

Query Match          1.8%; Score 143; DB 2; Length 536;
Best Local Similarity 33.8%; Pred. No.2,49e-04;
Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;

Db 139 OKKNHNLIRRRRNFINDRKEIETGLIPKSSPDEKRNKGTILKASVDYIRKLQEQOR 198
      ::: |||||::: |||||::: ||::: ||::: ||::: ||::: ||::: ||
OY 335 ERRRAHMLIEKKTRCSINDRIQLKVL--CGD-EAKLSKSTLRRALEHIEVHEHENV 411
      |::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||

OY 199 SKD-LESQRSLEQANR 214
      |::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
OY 412 LKHVEQMRKTL-QNNR 427

RESULT 14
ENTRY     S08343      #type complete
TITLE     nodulin precursor - soybean
ORGANISM  Glycine max #common_name soybean
DATE      29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change
                20-Aug-1999
ACCESSIONS S08343; S08344; A27059
REFERENCE   S08343
#authors    Franssen, H.J.; Thompson, D.V.; Idler, K.; Kormelink, R.; van
            Kamen, A.; Bisseling, T.
#journal    Plant Mol. Biol. (1989) 14:103-106
#title      Nucleotide sequence of two soybean ENOD2 early nodulin genes
            encoding Ngm-75.
#cross-references MUID:91332483
#accession  S08343
##molecule_type DNA
##residues  1-309 ##label FRA
##cross-references EMBL:X16876; NID:g18575; PIDN:CAA34758.1; PID:g18576
#experimental_source strain Wayne
#genetics    CH1
#accession   S08344
##molecule_type DNA
##residues  1-309 ##label FRZ
##cross-references EMBL:X16876; NID:g18578; PIDN:CAA34759.1; PID:g18579
#experimental_source strain Wayne
#genetics    CH2
REFERENCE     A94164
#authors      Franssen, H.J.; Nap, J.P.; Gloudeemans, T.; Striekema, W.; van
            Dam, H.; Govers, F.; Louwerse, J.; van Kamen, A.;
            Bisseling, T.
#journal      Proc. Natl. Acad. Sci. U.S.A. (1987) 84:4495-4499
#title        Characterization of cDNA for nodulin-75 of soybean: a gene
            product involved in early stages of root nodule
            development.

```

```
#accession A27059
#molecule_type mRNA
#residues 69-309 ##label FRW
GENETICS CH1
#gene ENOD2A
GENETICS CH2
#gene ENOD2B
CLASSIFICATION #superfamily proline-rich protein 3
KEYWORDS nitrogen fixation; nodulation
FEATURE 1-25
SUMMARY 26-309
#domain signal sequence #status predicted #label SIG
#product nodulin #status predicted #label MAT
SUMMARY #length 309 #molecular-weight 35966 #checksum 8577

Query Match 1.7%; Score 133; DB 2; Length 309;
Best Local Similarity 28.2%; Pred. No. 5,16e-03;
Matches 20; Conservative 17; Mismatches 33; Indels 1; Gaps 1;

Db 89 YLPPEKPPPEYQPPHEKPPHENPHEHQPPHEKPPPEYE-PPHEKPPPE 147
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 YMSPYDAQGSPGSPYSQHOSPPPHHHHPKIHENPEQYASPSIEDAPETKPTHL 263
: : : : : : : : : : : : : : : : : : : : : : : : :

Db 148 YQPPHEKPPPE 158
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 VEPQSPKSPQN 274

RESULT 15
ENTRY A36289 #type complete
TITLE CewYOD protein 1 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 24-Sep-1999
ACCESSIONS A36289; T15319
REFERENCE A36289
#authors Krause, M.; Fire, A.; Harrison, S.W.; Priess, J.; Weintraub, H.
#journal Cell (1990) 63:907-919
#title CewYOD accumulation defines the body wall muscle cell fate during Caenorhabditis elegans embryogenesis.
#cross-references MURD:91077929
#accession A36289
#status preliminary
#molecule_type DNA
#residues 1-324 ##label KRA
#cross-references GB:M59940; GB:M7497; NID:g455009; PID:g156387
#note the authors translated the codon TTG for residue 262 as Lys

REFERENCE Z18329
#authors Geisler, C.
#submission submitted to the EMBL Data Library, October 1995
#description The sequence of C. elegans cosmid B0304.
#accession T15319
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-324 ##label GEI
#cross-references EMBL:U39472; NID:g1041877; PID:g1041878; PIDN:AAA80133.1

GENETICS ##experimental_source strain Bristol N2
#gene h1h-1
#introns 577/2; 150/1; 241/3; 277/1; 297/1
KEYWORDS DNA binding; nucleus; transcription regulation
SUMMARY #length 324 #molecular-weight 36449 #checksum 5119

Query Match 1.7%; Score 137; DB 2; Length 324;
Best Local Similarity 22.7%; Pred. No. 1,36e-03;
Matches 29; Conservative 38; Mismatches 52; Indels 9; Gaps 7;

Db 111 KSTPNATELLIQSRVDSQHEDTTSTAGAGVG-GPRRTKPVLSVDRKKAATMERRRLR 169
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 RSKKAKAKIYIQETABGDEDEDESDSGETMSQGTITIVRRPKTERTAHNLIEKK-YR 368
```

```
Db 170 -KVNFAFEVVKQRTCPNPNOFLPRVEILLRSAIDYINLIER---ML-QOAGKMTIMEONQ 224
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 CSINDRIQQLKVLICGD-EAKLSKATLRRALIEHIEVEHENQYLKHHVEQMRKTL-QNN 426
: : : : : : : : : : : : : : : : : : : : : : : : :

Db 225 HLOMTQOI 232
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 RLPPPEPI 434
```

Search completed: Sun Sep 3 12:32:39 2000  
Job time : 103 secs

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
NWSEI  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Sep 3 12:28:01 2000; Maspar time 80.17 Seconds

Tabular output not generated. 962.525 Million cell updates/sec

Title: >US-09-332-522B-2

Description: (1-1113) from US09332522B.pep

Perfect Score: 8012

Sequence: 1 MNEEFEGDVPMSDPFLSLVT.....AQPDAFHLLTLVKLHTSMDL 1113

Scoring table: PAM 150

Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 54.388; Variance 102.303; scale 0.532

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	7998	99.8	1113	5	Q9XX00	0.00e+00
2	322	4.0	1113	5	Q24146	1.18e-36
3	265	3.3	403	11	Q9WNT3	7.73e-26
4	229	2.9	144	5	Q97676	2.80e-19
5	161	2.0	498	5	P91527	8.56e-08
6	149	1.9	110	11	Q88368	6.03e-06
7	149	1.9	377	11	Q70241	6.03e-06
8	149	1.9	419	4	Q14841	6.03e-06
9	149	1.9	468	13	Q73871	6.03e-06
10	149	1.9	474	11	Q08874	6.03e-06
11	149	1.9	520	4	Q75030	6.03e-06
12	151	1.9	1829	11	Q35889	3.00e-06
13	139	1.7	281	5	Q21653	1.83e-04
14	135	1.7	317	11	Q9WNT4	6.92e-04
15	140	1.7	320	10	Q92UV8	1.31e-04
16	136	1.7	347	4	Q14948	4.98e-04
17	137	1.7	567	5	Q15817	3.57e-04
18	125	1.6	223	5	Q9XZ11	1.73e-02
19	126	1.6	310	11	Q35410	1.26e-02
20	126	1.6	310	11	Q09135	1.26e-02

21	127	1.6	338	4	Q60409	9.19e-03
22	129	1.6	380	5	Q94749	4.86e-03
23	130	1.6	391	10	P93667	3.53e-03
24	131	1.6	405	10	Q03462	2.55e-03
25	131	1.6	410	10	Q41786	2.55e-03
26	126	1.6	415	10	Q23192	1.26e-02
27	127	1.6	417	5	Q02080	9.19e-03
28	132	1.6	434	3	Q12398	1.85e-03
29	126	1.6	524	10	Q80536	1.26e-02
30	129	1.6	604	13	Q42472	4.86e-03
31	126	1.6	675	5	Q61310	1.26e-02
32	131	1.6	1212	5	Q24523	2.55e-02
33	132	1.6	1816	4	Q75088	1.85e-03
34	124	1.5	219	13	Q91151	2.36e-02
35	124	1.5	405	10	Q41757	2.36e-02
36	121	1.5	509	13	Q91236	5.97e-02
37	123	1.5	592	5	Q9XV50	3.22e-02
38	124	1.5	790	10	Q23068	2.36e-02
39	121	1.5	876	2	Q45458	5.97e-02
40	122	1.5	974	10	Q49634	4.39e-02
41	120	1.5	1004	2	P71719	8.11e-02
42	124	1.5	1267	5	Q93564	2.36e-02
43	123	1.5	1271	5	Q21789	3.22e-02
44	124	1.5	1743	4	Q75087	2.36e-02
45	123	1.5	3164	14	Q69088	3.22e-02

## ALIGNMENTS

RESULT 1  
ID Q9XX00 PRELIMINARY; PRT; 1113 AA.

AC Q9XX00;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE Y47D3B.7 PROTEIN.

GN Y47D3B.7.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditiina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 94150718.

RA MATTHEWS L.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BOSEFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DO Z., DORBIN R., FAYELLO A., FULTON L., GARDNER A., GREEN S., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SAMPSON N., SMITH A., SONNHAMMER E., STRADEN A., SULLSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOHLDMAN P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Nature 368:32-38(1994).

DR EMBL: AL031635; CAA21042.1; -

SO SEQUENCE 1113 AA; 125156 MW; BA5E6E60 CRC32;

Query Match 99.8%; Score 7998; DB 5; Length 1113;

Best Local Similarity 99.6%; Pred. No. 0.00e+00;

Matches 1109; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db	1	MNEEFEGDVPMSDPFLSLVT	KLDDIAPPNNPDLDFMEHNMWQEPGSGQDPSPISGNH	60
Qy	1	MNEEFEGDVPMSDPFLSLVT	KLDIDAPPNNPDLDFMEHNMWQEPGSGQDPSPISGNH	60

```

Db 61 SPPEYDIDGORDVSTLHSLNHNDDFFSMRFPSPNFDLGGGRPSLAATQOLSGEAP 120
QY 61 SPPEYDIDGORDVSTLHSLNHNDDFFSMRFPSPNFDLGGGRPSLAATQOLSGEAP 120
Db 121 ASMLNPLOTSPSGGPPADATRPPLSLAQQALAPAMTPHQASLRYNIGIDOKNTAM 180
QY 121 ASMLNPLOTSPSGGPPADATRPPLSLAQQALAPAMTPHQASLRYNIGIDOKNTAM 180
Db 121 ASMLNPLOTSPSGGPPADATRPPLSLAQQALAPAMTPHQASLRYNIGIDOKNTAM 180
QY 121 ASMLNPLOTSPSGGPPADATRPPLSLAQQALAPAMTPHQASLRYNIGIDOKNTAM 180
Db 181 LSPPHHSTPQPYTEAMHNGVMSPYOAGGSPSYSOHOSPPHHHHPMPMT 240
QY 181 LSPPHHSTPQPYTEAMHNGVMSPYOAGGSPSYSOHOSPPHHHHPMPMT 240
Db 181 LSPPHHSTPQPYTEAMHNGVMSPYOAGGSPSYSOHOSPPHHHHPMPMT 240
QY 181 LSPPHHSTPQPYTEAMHNGVMSPYOAGGSPSYSOHOSPPHHHHPMPMT 240
Db 241 HENPEOVASPSIEDAPETKPTLVEPOSQPMKEELLRLVYNSPSEVERLKNKKS 300
QY 241 HENPEOVASPSIEDAPETKPTLVEPOSQPMKEELLRLVYNSPSEVERLKNKKS 300
Db 241 HENPEOVASPSIEDAPETKPTLVEPOSQPMKEELLRLVYNSPSEVERLKNKKS 300
QY 241 HENPEOVASPSIEDAPETKPTLVEPOSQPMKEELLRLVYNSPSEVERLKNKKS 300
Db 301 ACSATNGPSRSKKAAYIOETAGDEDEDESDSGTMSOGTTIYRRKTERTAH 360
QY 301 ACSATNGPSRSKKAAYIOETAGDEDEDESDSGTMSOGTTIYRRKTERTAH 360
Db 301 ACSATNGPSRSKKAAYIOETAGDEDEDESDSGTMSOGTTIYRRKTERTAH 360
QY 301 ACSATNGPSRSKKAAYIOETAGDEDEDESDSGTMSOGTTIYRRKTERTAH 360
Db 361 NLEKKYCSINDRIQOLVLLCGDEAKLSKATLRALTEHEVEHNOVLKHHEQMR 420
QY 361 NLEKKYCSINDRIQOLVLLCGDEAKLSKATLRALTEHEVEHNOVLKHHEQMR 420
Db 421 KTLQNNRLPEPIQYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDGSSKYL 480
QY 421 KTLQNNRLPEPIQYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDGSSKYL 480
Db 421 KTLQNNRLPEPIQYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDGSSKYL 480
QY 421 KTLQNNRLPEPIQYTEXSARSPVSSPPRNERKSRMSTTPMKNGTDGSSKYL 480
Db 481 FAMLAVLIFNPGLAGSAIFSKAAAEAPLASPEHGVYIDPDGSTRTLFWGSIIN 540
QY 481 FAMLAVLIFNPGLAGSAIFSKAAAEAPLASPEHGVYIDPDGSTRTLFWGSIIN 540
Db 541 M5YVWENLMTIYVVKLLIHGDPVQDM5VSWQTFVTRKARLNSGNLKDQRF 600
QY 541 M5YVWENLMTIYVVKLLIHGDPVQDM5VSWQTFVTRKARLNSGNLKDQRF 600
Db 601 CECGLATLDRSLPSPGVDSVFSVWECVRHLNMLWIGRIYARRRSTKPVSVVCSHAO 660
QY 601 CECGLATLDRSLPSPGVDSVFSVWECVRHLNMLWIGRIYARRRSTKPVSVVCSHAO 660
Db 661 TAVLYHEIHQHLMTGTFNEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAYMAOI 720
QY 661 TAVLYHEIHQHLMTGTFNEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAYMAOI 720
Db 721 YTSASIOCLALPNLLAPFSGYFLRRARRHVARPEHVSHTLTFHPATRKFSADAR 780
QY 721 YTSASIOCLALPNLLAPFSGYFLRRARRHVARPEHVSHTLTFHPATRKFSADAR 780
Db 781 LEHVLSSKQOLRFSGFEDEQLSPILARTITLKYLLSKLYOELVGDEIFTKNERIL 840
QY 781 LEHVLSSKQOLRFSGFEDEQLSPILARTITLKYLLSKLYOELVGDEIFTKNERIL 840
Db 841 NONDRLEDVVDVSRLLVITSTOCALITLNKEDSAFGTMSISNGACCTWTHVLT 900
QY 841 NONDRLEDVVDVSRLLVITSTOCALITLNKEDSAFGTMSISNGACCTWTHVLT 900
Db 841 NONDRLEDVVDVSRLLVITSTOCALITLNKEDSAFGTMSISNGACCTWTHVLT 900
QY 841 NONDRLEDVVDVSRLLVITSTOCALITLNKEDSAFGTMSISNGACCTWTHVLT 900
Db 901 CGIYMSKNELAROHYSILRNCPKITLDNGLAVGHALCARICIDDRDPSKYQYVC 960
QY 901 CGIYMSKNELAROHYSILRNCPKITLDNGLAVGHALCARICIDDRDPSKYQYVC 960
Db 961 IHTKLSLESLRLESTSSRASGVVSGIOEGTRMAYEMINSLLDAMRSLFASKPYWTOS 1020
QY 961 IHTKLSLESLRLESTSSRASGVVSGIOEGTRMAYEMINSLLDAMRSLFASKPYWTOS 1020
Db 1021 FFGQSFSTLYOEAYNHAYALINGTRDCCWRLFYELTCRLKNGANQATWSGYRRVRSK 1080
QY 1021 FFGQSFSTLYOEAYNHAYALINGTRDCCWRLFYELTCRLKNGANQATWSGYRRVRSK 1080
Db 1081 MDVAVGRVSMRBSAOPDAFHILHTLVKLHTSMOL 1113
QY 1081 MDVAVGRVSMRBSAOPDAFHILHTLVKLHTSMOL 1113

```

```

RESULT 2
ID Q24146 PRELIMINARY; PRT: 1113 AA.
AC Q24146:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
DE HLH106.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON S.
RX MEDLINE: 96165543.
RA THEOPOLD U., EKEGREN S., HULTMARK D.;
RT "HLH106, a Drosophila transcription factor with similarity to the
RT vertebrate steroid responsive element binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1195-1199(1996).
DR EMBL: U38238; AAA97864.1; -.
DR HSSP: P36956; 1AM9.
DR FLYBASE: FBgn0015234; HLH106.
DR PFAM: PF00010; HLH; 1.
SQ SEQUENCE 1113 AA; 124513 MW; E0D92420 CRC32;

Query Match 4.0%; Score 322; DB 5; Length 1113;
Best Local Similarity 24.6%; Pred. NO. 1,188-36;
Matches 192; Conservative 204; Mismatches 305; Indels 81; Gaps 68;

Db 261 SPAMEVOGKVPINVOQKVEKRSANHAIERISINDKINELKNLVGEQAKLMS 320
QY 337 SGETMS-OGTIIYV-RPKT-E-RTAHNLEKKYCSINDRIQOLKVLGDEAKLSKS 392
Db 332 AVLRKSIDKINDIQONNDLKAELQRLQRELMARNGSVKLLQGTGPRASKRRRESS 380
QY 333 ATLRAIEHIEVEHENOVK---HVE-Q-M-RK-T-LQNN-RLPY-P--EPIQYETS 440
Db 381 CFTTDACTLPSPDESPTSIPMHSDISLPPSPYGSSTACSSGSSSSNEEPLVPPSSM 440
QY 441 ASRPVSSPSPPR-NEKRS--RM-STTT-PMKN-CTRDS-SKVTLPFAMLAVALIFMPI 493
Db 441 RGMATHSRLGLCMFALAVNPKEFLQRGHYSDNDLGMGSGORILSTDSGEFAY 500
QY 494 -GLLA-GS-AI--FSKAA-AEAPIASPEHGVYIDPDG-GR-ST--RTL-FW-EG-ST-I 539
Db 501 WQSSWMLNLTMLTGLGVLKLYGDP-Q-LDAQT-LAYCOHROAIFYSOGSSQAY 557
QY 540 -NMS-YVWVFNLMTIYVVKLLIHGDPVQDM5VSWQTFVTRKARLNSGNLKDQ 597
Db 558 AGYLCLMHFGSLSPASRLCYLQTTWQFLRFLRWLGRVLSRSGGLFSMAARKQA 617
QY 598 RKFCGLATLDRSLPSPGVDSVFSVWECVRHLNMLWIGRIYARRRSTKPVSVCS 657
Db 618 LASARELALFNRLQLOLTGN--GS--RGDMNGIMALLFASNAEVA--HNLLPRETI 671
QY 658 HAQTAIVLHEIHQHLMTGTFNEDTPEPSALTGLFMSLCAYNLAEAGASNDGLPRAYM 717
Db 672 C-IHMTALBMKRSAPKQLOFFARYMSRAROEGRFRAEOT-QELRMAFTAYGYRY- 728
QY 718 AOIYTSASIOCLALPNLLAPFSGYFLRRARRHVARPEHVSHTLTFHPATRKFM 775
Db 729 CATHTVFTYDLSGBODEFTRLRN-PCDPAHVITKQYREHLFFKSIQCLVGAGHSGGL 787
QY 776 SDARKLEHVLSSKQOLRFSGFEDEQLSPILARTITLKYLLSKLYOELVG-GEIFTK 834
Db 788 PTSSVSGEAEOLQOQ-Q--HSGTIVSNV-LKYTSL-LKD-LMAD-ED--ERDTNV--VW 836
QY 835 INERTILNDRLEDVVDVSRLLVITSTOCALITLNKEDSAFGTMSISNGACCTW 894
Db 837 WADVLETAIVHMLGSEDTLAEQYGRIKQMPLOQCCGNDHLPRALHILAKMILKNN 896
QY 895 WTHVLTGCIYMSKNELAROHYSILRNCPKIT--TUNGLAVG-HA-LCARKICIDDR 950

```

RESULT	3	PRELIMINARY:	PRT:	403 AA.
ID	OSNTN3			
AC	OSNTN3:			
DT	01-NOV-1999 (TREMblrel. 12, Created)			
DT	01-NOV-1999 (TREMblrel. 12, Last sequence update)			
DT	01-NOV-1999 (TREMblrel. 12, Last annotation update)			
DE	STEROL REGULATORY ELEMENT-BINDING PROTEIN-1 (SREBP-1) (FRAGMENT).			
GN	SREBP-1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 99161303.			
RA	INoue J., SATO R.;			
RT	"A novel splicing isoform of mouse sterol regulatory element-binding			
RT	protein-1 (SREBP-1)."			
RL	Biosci. Biotechnol. Biochem. 63:243-245(1999).			
DR	EMBL: AB017337; BAA74795.1; -.			
FT	NON_TER	1		
FT	NON_TER	403	403	
SO	SEQUENCE	403 AA;	41018 MW;	BB57AA8F CRC32:
Query Match		3.3%.	Score 265;	DB 11;
Best Local Similarity		50.0%.	Pred. No. 7.72e-26;	
Matches	39;	Conservative	18;	Mismatches 21;
				Indels 0;
				Gaps 0

Db	272	OSREKTRAHNAIEKRRPSSINDKIVELKDIWGTAKLNKSAVLRAIDYIRFLQHSNQ	331
Oy	351	RPKTRERRAHNLIEKKYRCISINDBIQOKVLCDCEAKLSAFLRRAIEHIEVEHENQ	410
Db	332	KLKQENLITRSAAHKSLSL	349
Oy	411	VLKHHVEOMRKTIONNRL	428
RESULT	4	PRELIMINARY:	PRI: 144 AA.
ID	097576		
AC	097576:		
DT	01-MAY-1999	(TfEMBLrel. 10, Created)	
DT	01-MAY-1999	(TfEMBLrel. 10, Last sequence update)	
DT	01-NOV-1999	(TfEMBLrel. 12, Last annotation update)	
DE	ADIPCYOTE DETERMINATION AND DIFFERENTIATION-DEPENDENT FACTOR 1		
DE	(FRAGMENT).		
GN	ADP1.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Cetartiodactyla; Suidae; Sus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	DING S.T., MERSMANN H.J.;		
RL	Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL: AF102873; AAC78685.1; -		
DR	HSSP: P36956; IAM9.		
FT	NON TER	1	
FT	NON TER	144	144
SEQUENCE	144 AA: 15222 MW: 5486A656 CRC32:		
Query Match	2.9%: Score 229: DB 6: Length 144:		

```

Best local similarity 55.9%. Pred. No. 2.80e-19
Matches 33; Conservative 13; Mismatches 13; Indels 0; Gaps 0.

Db      85 GRALSGOSKQEKRTAHNAIEKRYRSSINDKIILDKLVGTEAKLNKSAYLRKAIDYI 143
       |:: ::||| ||| ||| ||| :||| :||| ||| ||| ||| :||| :||| :|||
Oy     344 GTTIVVRPTERTANTLLEKKRCISINDRIQLVLLCGDEAKLSKSTLRAAIHI 402

```

RESULT	5	PRELIMINARY:	PRT:	498 AA.
ID	P91527			
AC	P91527;			
DT	01-MAY-1997 (TrEMBLrel. 03, Created)			
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)			
DE	SMILABILITY TO A HELIX-LOOP-HELIX MOTIF.			
GN	W02C12.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;			
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE, 94150718.			
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,			
RA	BONFIELD J., BUTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,			
RA	CRAWFORD M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,			
RA	JONES M., KERSEW J., KIRSTEN T., LAISTER N., LAURELLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIEKEN L., ROOPER A., SAUNDERS D., SHAWKHEEN R.,			

```

RA INTERACTING WITH HOMAS R., WOLDMAN P., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994) .
[2]
RN SEQUENCE FROM H.A.
RC STRAIN-BRISTOL N2;
RA MURRAY J., WOLDMANN P.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U08015; AAB37997.1; -.
DR HSSP; P36956; IAM9.
DR PFAM; PF00010; HLH; 1.
SQ SEQUENCE 498 AA; 54391 MW; EAEBA806 CRC32;

Query Match          2.0%; Score 161; DB 5; Length 498;
Best Local Similarity 29.0%; Pred. NO. 8,56e-08;
Matches 29; Conservative 34; Mismatches 30; Indels 7; Gaps 6;

Db 238 KQDIHNMERRRNRNNDRIKELOGLPKNTSEDKLNKGITLKASCDYIFVLOKDREA 297
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 356 RRTAHNLIEKKYRKCSINDRIQOL-KVLL--CGDAKSKSKSLTKRALHLEVEHE-NQV 411
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 298 MKTQ-QQ-QKSLESTAHKYADRYKLEEMLARQGVOVPPS 335
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 412 LKHIVEGQRKTLONNRLPYPEPI-QYTEYSARSEVSPPS 450

RESULT      6
ID 008368    AC 008368;             PRELIMINARY;           PRT;       110 AA.

DT 01-NOV-1998 (TReMbRel_08, Created)
DT 01-NOV-1998 (TReMbRel_08, last sequence update)
DT 01-NOV-1999 (TReMbRel_12, last annotation update)
DE MICROHTALMAIA ASSOCIATED TRANSCRIPTION FACTOR (FRAGMENT).
CN MICROHTALMAIA
OS Rattus norvegicus (Rat).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RX MEDLINE: 98149870.
RA WEILBACHER K.N., HERSHEY C.L., TAKEMOTO C.M., HORSTMANN M.A.,
RA HEMESATH T.J., TASHJIAN A.H., FISHER D.E.;
RT "Age-resolving osteopetrosis: a rat model implicating microphthalmia
RT and the related transcription factor TFE3."
RL J. Exp. Med. 187:775-785(1998).
DR EMBL: AF029886; AAC26170.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 110
FT SEQUENCE 110 AA: 13155 MW; 7341AD76 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 110;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 20 OKKDNNLIERRRRINNDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 79
OY 355 ERRTAHNLIEKKYRCISINDRIQOLKVLCC-GDEAKL--SKSATLRAIEHIEVEHENOV 411
DB 80 AKD-LENROKLEHANR 95
OY 412 LKHNYEQMKRTION-NR 427

RESULT 7
ID 070241 PRELIMINARY; PRT: 377 AA.
AC 070241;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MTF (FRAGMENT).
GN WH.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98167856.
RA HODGKINSON C.A., NAKAYAMA A., LI H., SWENSON L.-B., OPDECAMP K.,
RA ASHER J.A. JR., ARNHEITER H., GLASER T.;
RT "Mutation at the anophthalmic white locus in Syrian hamsters:
RT haploinsufficiency in the Mltf gene mimics human Waardenburg syndrome
RT type 2."
RL Hum. Mol. Genet. 7:703-708(1998).
DR EMBL: AF020900; AAC15952.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 1
FT SEQUENCE 377 AA: 41728 MW; 83722925 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 377;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 162 OKKDNNLIERRRRINNDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 221
OY 355 ERRTAHNLIEKKYRCISINDRIQOLKVLCC-GDEAKL--SKSATLRAIEHIEVEHENOV 411
DB 222 AKD-LENROKLEHANR 237
OY 412 LKHNYEQMKRTION-NR 427

RESULT 8
ID 014841 PRELIMINARY; PRT: 419 AA.
AC 014841;
```

```
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MTF PROTEIN (MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR).
GN MTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKIN;
RX MEDLINE: 94348499.
RA TACHIBANA M., PEREZ-JORDA L.A., NAKAYAMA A., HODGKINSON C.A., LI X.,
RA SCHNEIDER M., MIKI T., FEX J., FRANCKE U., ARNHEITER H.;
RT "Cloning of MTF, the human homolog of the mouse microphthalmia gene
RT and assignment to chromosome 3p14.1-p12.3."
RL Hum. Mol. Genet. 3:553-557(1994).
RN [2]
RP SEQUENCE OF 1-11 FROM N.A.
RX MEDLINE: 98160190.
RA WATANABE A., TAKEDA K., PLOPLIS B., TACHIBANA M.;
RT "Epistatic relationship between Waardenburg syndrome genes MTF and
RT PAX3."
RL Nat. Genet. 18:283-286(1998).
DR EMBL: Z29678; CA82775.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
KM DNA-binding.
SQ SEQUENCE 419 AA: 46938 MW; 5CE200FD CRC32;

Query Match 1.9%; Score 149; DB 4; Length 419;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 204 OKKDNNLIERRRRINNDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 263
OY 355 ERRTAHNLIEKKYRCISINDRIQOLKVLCC-GDEAKL--SKSATLRAIEHIEVEHENOV 411
DB 264 AKD-LENROKLEHANR 279
OY 412 LKHNYEQMKRTION-NR 427

RESULT 9
ID 073871 PRELIMINARY; PRT: 468 AA.
AC 073871;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MTF.
GN CM19.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LECHORN; TISSUE-RETINAL PIGMENTED EPITHELIUM;
RX MEDLINE: 98133990.
RA MOCUIT M., MAZAKI Y., MIZUNO N., HAYASHI H., EGUCHI G.;
RT "Role of Mtf in differentiation and transdifferentiation of chicken
RT pigmented epithelial cell."
RL Dev. Biol. 193:47-62(1998).
DR EMBL: D88363; BAA25648.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
FT NON_TER 1 1
FT SEQUENCE 468 AA: 52416 MW; 0686639D CRC32;

Query Match 1.9%; Score 149; DB 13; Length 468;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

DB 253 OKKDNNLIERRRRINNDRIKELGTLPKSNDDPMRNKGTILKASVDYIRKLOREQOR 312
```



```

OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 313 TKE-LENROKLEHANR 328
OY 412 LKHVHEQMRKTLON-NR 427

RESULT 10
ID 008874 PRELIMINARY: PRT: 474 AA.
AC 008874: 008843: 008885: 060781: 060782:
DT 01-JUN-1998 (TReMBLrel. 05, Created)
DT 01-JUN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR
DE (PUTATIVE TRANSCRIPTION FACTOR MI) (MICROPHthalmia-RELATED PROTEIN).
GN MITF OR MI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=LIVER:
RX MEDLINE: 93345026.
RA HODGKINSON C.A., MOORE K.J., NAKAYAMA A., STEINGRIMSSON E.,
RA COPELAND N.G., JENKINS N.A., ARNHEITER H.;
RT "Mutations at the mouse microphthalmia locus are associated with
RT defects in a gene encoding a novel basic-helix-loop-helix-zipper
RT protein."
RL Cell 74:395-404(1993).
RN [2]
RP SEQUENCE OF 238-285 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=HEART;
RX MEDLINE: 94012591.
RA HUGHES M.J., LINGREL J.B., KRAKOVSKY J.M., ANDERSON K.P.;
RT "A helix-loop-helix transcription factor-like gene is located at the
RT mi locus."
RL J. Biol. Chem. 268:20687-20690(1993).
RN [3]
RP SEQUENCE OF 1-71 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=C57BL/6; TISSUE=HEART;
RX MEDLINE: 9519171.
RA STEINGRIMSSON E., MOORE K.J., LAMOREUX M.L., FERRE-D'AMARE A.R.,
RA BURLEY S.K., SANDERS ZIMRING D.C., SKOW L.C., HODGKINSON C.A.,
RA ARNHEITER H., COPELAND N.G., JENKINS N.A.;
RT "Molecular basis of mouse microphthalmia (mi) mutations helps explain
RT their developmental and phenotypic consequences."
RL Nat. Genet. 8:256-263(1994).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTION FACTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH TFE3 OR TFE8 (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN, ISOFORMS 1 AND 2,
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE
CC SEQUENCE SHOWN HERE IS THAT OF ISOFORM 1.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT HIGH LEVELS IN THE
CC HEART, SKIN AND SKELETAL MUSCLE. BARELY DETECTABLE IN OTHER
CC TISSUES. IN THE EMBRYO, EXPRESSED IN DEVELOPING EYE, EAR, SKIN AND
CC HEART.
CC -1- DISEASE: DEFECTS IN THE MITF GENE ARE THE CAUSE OF MICROPHthalmia,
CC A CONDITION WHICH IS CHARACTERIZED BY ONE OR MORE OF THE
CC FOLLOWING: LOSS OF PIGMENTATION, REDUCED EYE SIZE, FAILURE OF
CC SECONDARY BONE RESORPTION, REDUCED NUMBERS OF MAST CELLS AND EARLY
CC ONSET OF DEAFNESS, AND WHICH GIVES RISE TO A NUMBER OF DIFFERENT
CC PHENOTYPES.
CC -1- SIMILARITY: TO OTHER MEMBERS OF THE MYC FAMILY OF HELIX-LOOP-HELIX
CC TRANSCRIPTION FACTORS.
DR EMBL: L22958; AAB4773.1; -.
DR EMBL: Z23066; CA80600.1; -.
DR EMBL: U19874; AAC52155.1; -.
DR EMBL: U19875; AAC52156.1; -.
DR HSSP: P22415; IAN4.
DR MGD: MGI:104554; Mlrf.
DR PFAM: PF00010; HLH; 1.

```

```

KW Nuclear protein; Transcription regulation; Alternative splicing;
KW DNA-binding; Phosphorylation; Disease mutation.
FT DOMAIN 201 217 ARG/LYS-RICH (BASIC).
FT DOMAIN 267 288 LEUCINE-ZIPPER (POTENTIAL).
FT DNA BIND 218 260 HELIX-LOOP-HELIX MOTIF.
FT VARSPIC 1 66 MTSRLILROQLMRQMOERREDOQKLAQAOEFMOQVANS
FT OTPAINVSPTPLSPATOVEEVLK -> MLEMELISHYO
FT (IN ISOFORM 2).
FT VARIANT 67 202 MISSING (IN MICROPHthalmia WHITE SPOT).
FT VARIANT 243 246 MISSING (IN MICROPHthalmia SPOTTED).
FT VARIANT 242 267 ACIPRESEARALAKEROKDNHLL -> V (IN
FT MICROPHthalmia EYELSS-WHITE).
FT VARIANT 267 267 I -> N (IN MICROPHthalmia WHITE).
FT VARIANT 271 271 R -> K (IN MICROPHthalmia OAK RIDGE).
FT VARIANT 272 272 MISSING.
FT VARIANT 277 277 D -> N (IN MICROPHthalmia VITILIGO).
FT VARIANT 318 474 MISSING (IN MICROPHthalmia CLOUDY-EYED).
SQ SEQUENCE 474 AA; 52991 MW; 8E464430 CRC32;

Query Match 1.9%; Score 149; DB 11; Length 474;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

Db 259 OKDNHNIERRRPNDRIKELGTLPKSNDDPMRNKGTILKASYDITRKLOREQOR 318
OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 319 AKD-LENROKLEHANR 334
OY 412 LKHVHEQMRKTLON-NR 427

RESULT 11
ID 075030 PRELIMINARY: PRT: 520 AA.
AC 075030.
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE A-TYPE MICROPHthalmia ASSOCIATED TRANSCRIPTION FACTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 98321192.
RA AMAR S., FUSE N., YASUMOTO K., SATO S., YAJIMA I., YAMAMOTO H.,
RA UDENO T., DURLU Y.K., TAKAI M., TAKAHASHI K., SHIBAHARA S.;
RT "Identification of a novel isoform of microphthalmia-associated
RT transcription factor that is enriched in retinal pigment epithelium."
RL Biochem. Biophys. Res. Commun. 247:710-715(1998).
DR EMBL: AB006909; BA032288.1; -.
DR HSSP: P22415; IAN4.
DR PFAM: PF00010; HLH; 1.
SQ SEQUENCE 520 AA; 58162 MW; F5D32118 CRC32;

Query Match 1.9%; Score 149; DB 4; Length 520;
Best Local Similarity 31.2%; Pred. No. 6.03e-06;
Matches 24; Conservative 27; Mismatches 21; Indels 5; Gaps 4;

Db 305 OKDNHNIERRRPNDRIKELGTLPKSNDDPMRNKGTILKASYDITRKLOREQOR 364
OY 335 ERRTAHNIIEKKYCSINDRIQOLKVLIC-GDEAKL--SKSATLRALIEHIEVEHENOV 411
Db 365 AKE-LENROKLEHANR 380
OY 412 LKHVHEQMRKTLON-NR 427

RESULT 12
ID 035889 PRELIMINARY: PRT: 1829 AA.
AC 035889.
DT 01-JAN-1998 (TReMBLrel. 05, Created)

```

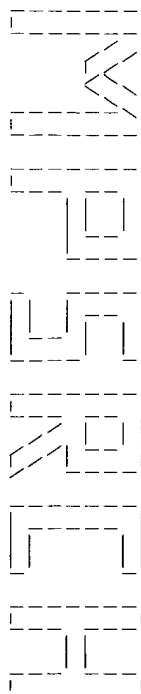
elegrans."

DT	01-MAY-1999 (TREMBlre1. 10, Last sequence update)
DT	01-MAY-1999 (TREMBlre1. 10, Last annotation update)



480 251 / 160  
09

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sun Sep 3 12:26:41 2000; Maspar time 34.32 Seconds  
Tabular output not generated. 1005.084 Million cell updates/sec

Title: >US-09-332-522B-2  
Description: (1-1113) from US09332522B.pep  
Perfect Score: 8012  
Sequence: 1 MNEFEQDVPMSPDFLSLVY.....AQPDAFHLLTLVLTSLTMDL 1113

Scoring table: PAM 150  
Gap 11

Searched: 95661 seqs, 30989116 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 56.369; Variance 100.483; scale 0.561

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	443	5.5	1139	1	SRE2_CRIGR	STEROL REGULATORY ELEM
2	421	5.3	1141	1	SRE2_HUMAN	STEROL REGULATORY ELEM
3	387	4.8	1133	1	SRE1_CRIGR	STEROL REGULATORY ELEM
4	375	4.7	1023	1	SRE1_RAT	STEROL REGULATORY ELEM
5	358	4.5	1147	1	SRE1_HUMAN	STEROL REGULATORY ELEM
6	143	1.8	446	1	TFE3_MOUSE	TRANSSCRIPTION FACTOR E
7	145	1.8	477	1	TFE3_HUMAN	TRANSSCRIPTION FACTOR E
8	143	1.8	743	1	TFE3_HUMAN	TRANSSCRIPTION FACTOR E
9	133	1.7	309	1	NO75_SOYB	EARLY NODULIN 75 PRECU
10	137	1.7	324	1	MYOD_CAEEL	MYOBLAST DETERMINATION
11	133	1.7	329	1	MYOD_CAEEL	MYOBLAST DETERMINATION
12	140	1.7	413	1	ESCL1_SCHPO	ESCL1 PROTEIN.
13	128	1.6	260	1	SGM1_LYTV	TRANSCRIPTION FACTOR S
14	126	1.6	310	1	USF1_HUMAN	UPSTREAM STIMULATORY F
15	126	1.6	310	1	USF1_RABIT	UPSTREAM STIMULATORY F
16	126	1.6	310	1	USF1_MOUSE	UPSTREAM STIMULATORY F
17	127	1.6	387	1	TAP4_HUMAN	TRANSFORMING PROTEIN Q
18	131	1.6	387	1	QIN_AYV53	TRANSFORMING PROTEIN Q
19	132	1.6	451	1	BF1_CHICK	TRANSFORMING PROTEIN Q
20	127	1.6	514	1	TFEB_HUMAN	TEEB PROTEIN (FRAGMENT
21	130	1.6	1110	1	TRAI_CAEEL	SEX-DETERMINING TRANSF
22	125	1.6	1329	1	FTSX_ECOLI	CELL DIVISION PROTEIN
23	125	1.6	2248	1	CVI1_DROME	CA2+/CALMODULIN-RESPON

24	123	1.5	291	1	USF2_RAT	UPSTREAM STIMULATORY F
25	118	1.5	307	1	USF1_XENBO	UPSTREAM STIMULATORY F
26	123	1.5	346	1	USF2_MOUSE	UPSTREAM STIMULATORY F
27	123	1.5	346	1	USF2_HUMAN	UPSTREAM STIMULATORY F
28	117	1.5	352	1	NDP1_XENLA	NEUROGENIC DIFFERENTI
29	118	1.5	428	1	FKH4_MOUSE	TRANSCRIPTION FACTOR F
30	117	1.5	431	1	ACRO_RABIT	ACROSIN PRECURSOR (EC
31	120	1.5	458	1	YNE1_CAEEL	HYPOTHETICAL 52.8 KDA
32	124	1.5	480	1	BF1_RAT	TRANSCRIPTION FACTOR B
33	123	1.5	492	1	COX1_PHYME	CYTOCHROME C OXIDASE P
34	123	1.5	542	1	THSA_METH	PROBABLE THERMOSOME SU
35	119	1.5	2441	1	CAP_MOUSE	CREB-BINDING PROTEIN
36	123	1.5	3164	1	TEGU_HSV1	LARGE TEGUMENT PROTEIN
37	114	1.4	160	1	MAX_CHICK	MAX PROTEIN.
38	115	1.4	242	1	MYF6_HUMAN	MYOGENIC FACTOR MYF-6.
39	115	1.4	315	1	ATB3_XENLA	ATONAL PROTEIN HOMOLOG
40	114	1.4	459	1	STB1_YEAST	STB1 PROTEIN.
41	116	1.4	481	1	BF1_MOUSE	TRANSCRIPTION FACTOR B
42	115	1.4	514	1	MEPD_MOUSE	MYOCTE-SPECIFIC ENHAN
43	115	1.4	758	1	YM38_YEAST	HYPOTHETICAL 85.0 KDA
44	116	1.4	842	1	CLA4_YEAST	SERINE/THREONINE-PROTE
45	115	1.4	1169	1	SUV3_DROME	SUPPRESSOR OF VARIEGAT

## ALIGNMENTS

RESULT 1  
ID SRE2\_CRIGR STANDARD: PRT: 1139 AA.  
AC 060429: 060418, 060428: 060427;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE SRE2\_CRIGR ELEMENT-BINDING PROTEIN-2 (SREBP-2) (STEROL  
REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2).  
GN SREBP2 OR SREBP2.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
RN [1]  
RP SEQUENCE FROM H.A., AND TRUNCATED FORM SRD-1.  
RX MEDLINE: 95047343.  
RA Yang J., Sato R., Goldstein J.L., Brown M.S.;  
RT "Sterol-resistant transcription in CHO cells caused by gene  
rearrangement that truncates SREBP-2.";  
RL Genes Dev. 8:1910-1919(1994).  
RN [2]  
RP SEQUENCE FROM H.A. (TRUNCATED FORMS SRD-1 TO SRD-3).  
RX MEDLINE: 95263566.  
RA Yang J., Brown M.S., Ho Y.K., Goldstein J.L.;  
RT "Three different rearrangements in a single intron truncate sterol  
regulatory element binding protein-2 and produce sterol-resistant  
phenotype in three cell lines. Role of introns in protein  
evolution.";  
RL J. Biol. Chem. 270:12152-12161(1995).  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL  
REGULATORY ELEMENT 1 (SRE-1) (5'-ATACCCAC-3') FOUND IN THE  
FLANKING REGION OF THE LDLR AND HMG-COA SYNTHASE GENES (BY  
SIMILARITY).  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
BLH PROTEIN.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR  
ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS  
UPON PROTEOLYTIC CLEAVAGE.  
CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPs ARE PROTEOLYTICALLY  
CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS  
AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE  
PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,  
INDEPENDENT OF STEROL LEVELS.  
CC -1- DISEASE: STEROL-RESISTANT DEFECTIVE (SRD) PHENOTYPES EXPRESS  
TRUNCATED FORMS OF SREBP-2 PROTEIN, OFTEN FOUND FUSED TO OTHER  
PROTEINS, AS IS THE CASE IN SRD-1, WHERE SREBP-2 IS FUSED TO AN  
OUT-OF-FRAME KU P70 PROTEIN OR, IN SRD-2 WHERE THE FUSION PROTEIN



```

CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC INDEPENDENT OF STEROL LEVELS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02031; AAA50746.1; -.
DR HSSP: P36956; 1AM9.
DR PFM: 600481; -.
DR PFM: PF00010; HLH: 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE NEG.
KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW Endoplasmic reticulum.
FT DOMAIN 1 30 TRANSSCRIPTIONAL ACTIVATION (ACIDIC).
FT DOMAIN 51 124 GLY/PRO/SER-RICH.
FT DOMAIN 125 246 GLY/PRO/SER-RICH.
FT DOMAIN 95 423 BINDS TO STEROL REGULATORY ELEMENT-1 (BY
FT SIMILARITY).
FT DNA_BIND 330 343 BASIC DOMAIN.
FT DOMAIN 344 381 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 380 401 LEUCINE-ZIPPER (POTENTIAL).
FT SITE 468 469 CLEAVAGE (BY APOPAIN AND CASPASE-7).
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT DOMAIN 591 595 POLY-ALA.
SQ SEQUENCE 1141 AA: 123673 MW: DE42B9C16C832B CRC64:
Query Match 5.3%; Score 421; DB 1; Length 1141;
Best Local Similarity 25.1%; Pred. No. 5,596-59;
Matches 190; Conservative 210; Mismatches 280; Indels 77; Gaps 59;
Db 326 PKEBERRTHIIIEKRRSSINDKIELKLDVMTGDAKMKSGVIRKADIVYKIQVYVNH 385
QY 352 PK-TERRTAHMLIEKKYKCSINDRIQQLKVLICGDEAKLSKSLRAIEHT---EEVEH 407
Db 386 KLROENWMLKLANOKNKLKGLDGLVDNEVDKIEDFNQNVLLMSPASSSSQAQFS 445
QY 408 ----ENOVLK--HHVEQKRKTLQNNRLPYPE-PIQYTESASPSPESSPPRNERK--R 458
Db 446 PYSIDSESPFLDDAKYKDEPDSPVAGWDRSRILLVTLFELCSFNPPLTSLQMG 505
QY 459 S-RM-S-TTTPM-KNGT-RD--GSSKVTL-F--A-MLAVLIENPILLGSAIFSKAA 506
Db 506 AHSDDQHHSSGSGSVLSFESGSG-GWF-D-WMPPTLLMLVNGYVLSVFKLLHGE 562
QY 507 AEADIASFE-HGRVIDPDGTSTRTLLFWESIIIMSYVFNIMITYVVKLLIHSDP 565
Db 563 VIRBHSRSVFEWRRKQADLDLAGDFAAANQOTCLAVGRALPTSRDLASLSMN 622
QY 566 VQDFNSVSWQTFVITREARAEVLSNGNLKDAORKCECLATIDRSLPBGVDSFVSGME 625
Db 623 VIRSYLQKRLVRLMLLKVKVFCRRATPATEAGFEDEAKTSARDAALAYRLHQLH---IT 679
QY 626 CVRH-LIN-WL--WI-GR-YIARRRRSTTKPV-SYVCRSHQOTAVL-YHEIHQHLMIT 677
Db 680 KGLP-A-GSAC-SDVHMLCAVNAECA---EKEIPSTIVEIHLTAAMGLKTRCGKLG 733
QY 678 GNFEPTYPSALTLGFLMSLCVNLAEAGASNDGJPRAVMQIYISAIQCLALPNDLA 737
Db 734 -FLASYFLSPKOS-LC-GPEHSAVPDSLRLMLCHPLGOKFEMFRMSVYS-AKES-L-Y- 786
QY 738 PFFSGYFLRRARRHVRAPSHS-VSHLL-WIHPATRKFMDSAKLEHVLSSKOKOLRFG 795

```

```

Db 787 C-AQRNPADPIAQVHOAFCKNLERALTESLVKPOAK-KKAGDQ-EESCEFFSALEYL-- 841
QY 796 SEVEDQSLPARIITLLKVLKSLVQELVGGEIEFKNVERILINDRDLDDVDVY 855
Db 842 -KLHSP-VDSVGWSPPLSSVLKS--ALGPDLICRWMTSATVAIWSLQGDAAVRS 897
QY 856 SRLVITISTQCAALLTNKEDSAKFGWISNNGACCMTHVLTGCIYMSNKNELARQ 915
Db 898 HETKTERPKALEVETESPLYKAIFHA-CRAMHASPGRKADQOSSF-C-HCERA-SG-HL 952
QY 916 HYLIRNCPYKI-LTDN-LGLAVGHALC-AKICIDDRDSKVQYVCIFHKKSLEIRL 972
Db 953 WSSLNWSGSDPALNHYVQLLTQDLLL-SLRTALMOKQASQAV-GETYHASCAGELAG 1010
QY 973 FSTSRAGGVV-SGIQGETRMAVEWIMNSLLDA-WRSNLFASPYMTQSFQGGSTFTL 1030
Db 1011 FQDRLGSLRLAHSPRAYKRVFLHEATVRLMAGGSP 1047
QY 1031 YQEAYNHYA-IINGTRGDCWRLFVYELTCRMLNGANP 1066
RESULT 3
ID SREL_CRIGR STANDARD; PRT: 1133 AA.
AC Q60416;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-FEB-2000 (Rel. 39, last annotation update)
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL
DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).
GN SREBP1 OR SREBP1.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RX Sato R., Yang J., Wang X., Evans M.J., Ho Y.K., Goldstein J.L.,
RA Brown M.S.;
RT "Assignment of the membrane attachment, DNA binding, and
RT transcriptional activation domains of sterol regulatory
RT element-binding protein-1 (SREBP-1).";
RL J. Biol. Chem. 269:17267-17273(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL
CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCCGAC-3') FOUND IN THE
CC FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.
CC ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY, BINDING TO
CC BOTH AN E-BOX MOTIF (ATCACGTA) AND TO SRE-1.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE OF THE NUCLEAR
CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON
CC PROTEOLYTIC CLEAVAGE (BY SIMILARITY).
CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,
CC INDEPENDENT OF STEROL LEVELS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09103; AAA20085.1; -.
DR HSSP: P36956; 1AM9.
DR PFM: PF00010; HLH: 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.

```

Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;  
KM Endoplasmic reticulum; Activator; Apoptosis.  
FT DOMAIN 1 60 TRANSCRIPTIONAL ACTIVATION (ACIDIC).  
FT DOMAIN 61 176 TO STEROL/SER-RICH.  
FT BINDING 91 410 TO STEROL REGULATORY ELEMENT-1  
(PROBABLE).  
FT DNAS\_BIND 317 330 BASIC DOMAIN.  
FT DOMAIN 331 368 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 367 388 LEUCINE-ZIPPER (POTENTIAL).  
FT DOMAIN 421 454 GLY/PRO/SER-RICH.  
FT SITE 452 453 CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY  
SIMILARITY).  
FT TRANSMEM 477 497 POTENTIAL.  
FT TRANSMEM 536 556 POTENTIAL.  
SQ SEQUENCE 1133 AA; 120465 MW; 01A77B09DEDCDA84 CRC64;  
Query Match 4.8%; Score 387; DB 1; Length 1133;  
Best Local Similarity 29.4%; Pred. No. 8,25e-52;  
Matches 149; Conservative 132; Mismatches 17; Indels 54; Gaps 41;  
Db 313 QSRGERTAHNAIEKRRSSINDKIVELKDVTEAKLNKSAVLRAIDYIRIQHSNQ 372  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 351 RKTERBRIAHNIEKKRCSINDRIQQLKVLGDEAKLSKATLRRAIEHIEVEHEHQ 410  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 373 KIKQENLARNAAHKSLSKIDVSAAGSGGTDVAMEGYKPEYVDLTPPPSDAGSPSOS 432  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 411 VLKHEVEOMRKTLQNNR-L-PY-PE-PIQYTESARS--P--VES-SPSP-RERKRKS 459  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 433 SPLSGRSSSGSDSPEDSPVFEISOYKAQRLSHGMLDGRSLALCALVFLCTCNPL 492  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 460 R-MSTTTP-MKNGTRDGSCKTLFA-MLIAYLIENPIGLASAI-F-SKA-A-AE-API 511  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 493 ASLFGMGIPGSSASGAHSSGRSMLEAESRDGSMWTQWLPLPLVLANGLVLACLAL 552  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 512 ASPEFHG-RVIDPDGT--ST-RTLF---WEGSI-IN--MS-YWVVEIIMIIIVYVKL 559  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 553 FYVGEPTVRPHSPAVH-FWRHRKQADLDLNGDFAQAQOIMLQALGRPLPSNDL 611  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 560 LIHGPPV-ODEFVSVMQFVFTTRERAKRELNGNLKDAQRKCECLATLDRSLPSPGVDS 618  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 612 ACSLLMNLIRHLQRLWGRWLAGRAGGLRDCGLRMDARASARDAALVYHHLHOLHAMG 671  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 619 VFSVMECVRHILNLMWIGRIYARRRRSTTKPVSV-V-CRSAGQAVL-YHEIHQHLMG 675  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 672 --KX-T-G-GHLIASMLALSALMEACAG--DAVSMTAETIYVAALRVKTSLPRA 721  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 676 ITGNEDITYEPALTGLFWSLCAVNLAEAGASNDGLPRVAVAQIYISASICRLALPVL 735  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 722 L-HETREFLSAR-QACLAQSGVPLAMQWLCHPVGHFEPDGMVAVHG-AP-QESL-Y 776  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 736 LAFFSGYFLRRARRHVRAPRPHSVS-HLWTFHPATRKFMKDARLEHVLSSKQOLRF 794  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 777 -S-VAGNVDPDLAQTRLFCEHLER 800  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 795 GSFEDEQLSPARITRTILKYVLSK 820  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
RESULT 4  
ID SREL\_RAT STANDARD; PRT; 1023 AA.  
AC P56720;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL  
REGULATORY ELEMENT BINDING TRANSCRIPTION FACTOR 1) (ADIPOCYTE  
DETERMINATION AND DIFFERENTIATION-DEPENDENT FACTOR 1) (ADD1)  
(FRAGMENT).  
DE SREBP1 OR SREBP1.  
GN Rattus norvegicus (Rat).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
SEQUENCE FROM N.A.

TISSUE-ADIPOCYTE;  
RC MEDLINE; 93330269.  
RX Tontonoz P., Kim J.B., Graves R.A., Spiegelman B.M.;  
RA "ADD1: a novel helix-loop-helix transcription factor associated with  
RT adipocyte determination and differentiation.";  
RL Mol. Cell. Biol. 13:4753-4759(1993).  
RN [2]  
RP DNA-BINDING.  
RX MEDLINE; 95257939.  
RA Kim J.B., Spotts G.D., Halvorsen Y.D., Shih H.M., Ellenberger T.,  
RA Towle H.C., Spiegelman B.M.;  
RT "Dual DNA binding specificity of ADD1/SREBP1 controlled by a single  
RT amino acid in the basic helix-loop-helix domain.";  
RL Mol. Cell. Biol. 15:2582-2588(1995).  
CC -1- FUNCTION: REGULATED DURING BOTH ADIPOCYTE DETERMINATION AND  
CC DIFFERENTIATION. ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE  
CC SPECIFICITY. BINDING TO BOTH AN E-BOX MOTIF (ATCAGCGGA) AND TO  
CC SRE-1 (5'-ATCAGCCGC-3').  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC BHLH PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; SREBP-1A AND SREBP-1C/ADD1  
CC (SHOWN HERE) ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BROWN ADIPOSE  
CC TISSUE.  
CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY  
CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS  
CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE  
CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,  
CC INDEPENDENT OF STEROL LEVELS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. MANY PROBABLE  
CC FRAMESHIFTS WERE CORRECTED FROM POSITION 878 ONWARD.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, L16995; NOT\_ANNOTATED\_CDS.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
DR Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;  
KW Endoplasmic reticulum; Activator; Apoptosis; Alternative splicing.  
FT BINDING 67 385 TO STEROL REGULATORY ELEMENT-1 (BY  
FT SIMILARITY).  
FT DOMAIN 125 128 POLY-PRO.  
FT DNAS\_BIND 293 306 BASIC DOMAIN.  
FT DOMAIN 307 344 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 343 364 LEUCINE-ZIPPER (POTENTIAL).  
FT SITE 427 428 CLEAVAGE (BY APOPAIN AND CASPASE-7) (BY  
FT SIMILARITY).  
FT TRANSMEM 454 474 POTENTIAL.  
FT TRANSMEM 513 533 POTENTIAL.  
FT NON\_TER 1023 1023  
SQ SEQUENCE 1023 AA; 108571 MW; 2593DF46B2A11C8F CRC64;  
Query Match 4.7%; Score 375; DB 1; Length 1023;  
Best Local Similarity 30.2%; Pred. No. 2,65e-49;  
Matches 155; Conservative 125; Mismatches 180; Indels 53; Gaps 37;  
Db 282 GKALSGASREKRTAHNAIEKRRSSINDKIVELKDVTEAKLNKSAVLRAIDYIR 341  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 344 GTTIVRRPKRRAHNLIEKKRCSINDRIQQLKVLGDEAKLSKATLRRAIEHIE 403  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 342 FLQHSNOKLQKQENTLRSAAHKSLSKIDVSAAGSGGTDVAMEGYKPEYVDLTPPPSDA 401  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Qy 404 EVEHENVOLKHEVEOMRKTLQNN--R-LPY-PIQYTES--ARSP--VES-SPSP-R 453  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 402 GSPSGSPLSGRSSSGSDSPEDSPAFEDNOYKAQRLPSHGMIDPRLALCVLVF 461  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||



QY 454 NEKRSR-MSTTP-MKNGTRGSSKVTTLFA-MLL-AV-LIENPDLIAGS--A--IFS- 503  
 DB 462 LCLTCLNPLASLFGWGLITPSDASGVHRSRSGMYLEASRDGSMWTOMLPLVWLANGLL 521  
 QY 504 KAAAEPIASPFHGRVI-DDPGT--ST-RILE----WGST-IN--MS-YWVNIIM 551  
 DB 522 VLACLALLFVYGVPTRPHSGPAVHFRHRRKQADLDLARGDFAQAQOIMLALQALGRPL 581  
 QY 552 IIVYVVKLLIHGDPPVODFMSVSWQTFVTRKARAEINSNGLKDAORKECECIATLDRSL 611  
 DB 582 PHSNLDLACSLMLNIVRHLLQRLWGRWMLAGQAGGQRORYRLKKDARASRAADVAVYHKL 641  
 QY 612 PSPGVDSFVSQVCEVRLHMLWIGRIYARRRRSTTK--PVSVCGRSHAO-TAVLYHET 668  
 DB 642 HOLHMG--KYTGGH-LVA-SNL--ALSALNLAECAG--DAISMATLAEIYVAALRY 691  
 QY 669 HOLHMGITNGFEDTPEPSALTGLFMSLCVNLAEAGASNDGLPRAVYAOIITSISIQ 728  
 DB 692 KTSLPRAL-HLITREFLSSAR-QACLAQSGAVPLAMQWICHYPGARFVYDGDMAVHG-AP 748  
 QY 729 RLALPMLLAFPSFGYFLRRARRHVRAPHSVS-HLWIFHPRATRKREMSDARKLEHYLSS 787  
 DB 749 -QESL-Y-S-VAGNPVDPPLAQVTRLFCHELLER 777  
 QY 788 KOKQLRFGSFEDEQLSPLARIRITTLKXYLLSK 820  
 RESULT 5 SRE1 HUMAN STANDARD: PRT: 1147 AA.  
 AC P36956: 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DI 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE STEROL REGULATORY ELEMENT BINDING PROTEIN-1 (SREBP-1) (STEROL  
 REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 1).  
 GN SREBP1 OR SREBP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N. A., AND PARTIAL SEQUENCE.  
 RX MEDLINE: 94005541.  
 RA Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,  
 RA Goldstein J.L., Brown M.S.;  
 RT "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that  
 RT controls transcription of the low density lipoprotein receptor  
 RT gene".  
 RL Cell 75:187-197(1993).  
 RN [2]  
 RL X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 319-394.  
 RP MEDLINE: 98298442.  
 RA Partridge A., Bellisolell L., Ferre-D'Amaré A.R., Burley S.K.;  
 RT "Co-crystal structure of sterol regulatory element binding protein 1a  
 RT at 2.3-A resolution".  
 RL Structure 6:661-672(1998).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL  
 CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCAC-3'), FOUND IN THE  
 CC FLANKING REGION OF THE LDLR GENE AS WELL AS OTHER GENES.  
 CC ADD1/SREBP1 HETERODIMER HAS DUAL SEQUENCE SPECIFICITY. BINDING TO  
 CC BOTH AN E- BOX MOTIF (ATCAGTGA) AND TO SRE-1.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR  
 CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS UPON  
 CC PROTEOLYTIC CLEAVAGE.  
 CC -1- ALTERNATIVE PRODUCTS: SEVERAL ISOFORMS DIFFERING IN THE N- OR C-  
 CC TERMINI ARE PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN  
 CC IS THAT OF SREBP-1A.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES, MOST  
 CC ABUNDANT IN LIVER AND ADRENAL GLAND. IN FETAL TISSUES LUNG AND  
 CC LIVER SHOWS HIGHEST EXPRESSION.  
 CC -1- PIM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY  
 CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS

CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE  
 CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,  
 CC INDEPENDENT OF STEROL LEVELS.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 CC EMBL: U00968; CAB34682.1; -.  
 DR PDB: 1AM9; 01-JUL-98.  
 DR TRANSFAC; T01556; -.  
 DR MIM; 184756; -.  
 DR PFAM; PF00010; HLH; 1.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;  
 KM Endoplasmic reticulum; Activator; Apoptosis; 3d structure;  
 FT DOMAIN 61 178  
 FT BINDING 92 416  
 FT DNA\_BIND 323 336  
 FT DOMAIN 337 374  
 FT DOMAIN 373 394  
 FT DOMAIN 427 462  
 FT SITE 460 461  
 FT TRANSMEM 488 508  
 FT TRANSMEM 548 568  
 SQ SEQUENCE 1147 AA; 121644 MW; 58ED93A2371FE3E9 CRC64;  
 Query Match 4.5%; Score 358; DB 1; Length 1147;  
 Best Local Similarity 25.9%; Pred. No. 8,95e-46;  
 Matches 195; Conservative 195; Mismatches 299; Indels 72; Gaps 54;  
 Db 87 LSGPQAPSPPLPPAPPLPKMYSPMAFSPGPKIESVP-LSTLQTPPTPOPLGALL 145  
 QY 115 LSG-EGPASHLNPLOPSP-PSGGYPADAYRPLSLAQOLAAAMPHPQAAISFVNTGID 172  
 Db 146 POSFP-AP-APQFSESTPYLGYRSPPGCFSTSGSPGNTQQLPLGULPASPQVPSLHT 203  
 QY 173 QKNFTHMLTSPPHHTSMISQPYTEAMGHNTGYMSPYDQAQGSGBSYSQHHQSPPHH 232  
 Db 204 QVQSVVP-QCLLTVTAPRAAPVTTVTTSIQIOQVPLLOPHFIKADSL-LTAMK-TDGA 260  
 QY 233 HHHPMKIHENEOYASPSIEDAPETKPTHLVPEQSPKSPQMKKEILLRLNMSPEVE 292  
 Db 261 TVKA--AGISPLVSGTGYVGTGPIPLVSGGTLATVPLVDAEKLPINRLAASKAPASA 318  
 QY 293 RLKMKSCGASATNGPSRFRKAKAIVDETAEGBED-EDDESDSGETMSGGTTIVR- 350  
 Db 319 QSRGEKRTAHNAIEKRYBSINDKILKLDLVYGTEAKNKSAVLRAKADYTRFLOHSQ 378  
 QY 351 RKTERRTAHNIEKRYBSINDRLOQLVLLCGDEAKLSKATLRRALTEHIEVEHENO 410  
 Db 379 KIKQENLSRTVAHRSKSLKDLYVAGSGGNTDVMEGKTEVEPTLPPPDAGSPFGS 438  
 QY 411 VTKHVEQMKTLQNNR-L-PY-PE--PLYTE-Y--SKRSVESSPSPPRNE--R-KR 458  
 Db 439 SPLSLGSRGSGSGSDSEPDSPVFEDSKAPFEORPSLBSGMDRLSRALCTIVFLCL 498  
 QY 459 SRMS-TT-TPMNGT-ROSSKAVTLF--AMLAVALI--FNPGLLA-G--S--A-IF--- 502  
 Db 499 SCNPPLASLIGANGLPSPSOTTSVYHSPGRNVLTGTSRDGPGQAQMLLP-PVWLLGLIV 557  
 QY 503 S-KA-AA--EAP-IASPFHGRVIDDP--D--GTSTRT-LFEGSGIINMSYVWVNIIM 552  
 Db 558 LVSIVLFFYGGPVTRPHSGPAVHFRHRRKQADLDLARGDFAQAQOIMLALQALGRPL 617



Db	39	HNHSHRPHHHHHHHHHHHHHPPRPAPQPPRPAPAAQQOOPPPRPPLAPQAGAAQNSDEKGPQ	98
Oy	223	HHOSPPRRHHHHHHHPMK.IHNEDEOVASPIDADEETKPTLH.VPQSPKSPONKKELLR	2811
Db	99	LLL.LPPLDHHRRPSPGAAGCCRGCGELGPRVPEKEKGAAGAGEEKKGAEGSGKQDEGC	157
Oy	282	LVNNSPSEVERL.KNKKSGACS.ATN-GP-SRS-KEKAATVIOET.ABQDEDEDEDS	3355
Db	158	KEGEKN 164	
Oy	336	DSGETMS 342	
RESULT	8		
ID	TFE3_HUMAN	STANDARD:	PRT: 743 AA.
AC	P19532; Q99964; Q92757; Q92758;		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-FEB-2000 (Rel. 39, Last annotation update)		
DE	TRANSSCRIPTION FACTOR E3.		
GN	TFE3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RN	[1]		
RP	SEQUENCE OF 1-219 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RX	MEDLINE: 97140324.		
RA	Weerman M.A.J., Wilbrink M., Geurts van Kessel A.;		
RT	"Fusion of the transcription factor TFE3 gene to a novel gene, PRC3,		
RT	in t(X;1)(p11;q21)-positive papillary renal cell carcinomas."		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15294-15298(1996).		
RN	[2]		
RP	SEQUENCE OF 149-743 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RC	TISSUE=MONOCYTES.		
RX	MEDLINE: 97026295.		
RA	Sihar S.K., Clark J., Gill S., Hamoudi R., Crew A.J.,		
RA	Gwilliam R., Ross M., Linehan W.M., Birdsall S., Shipley J.,		
RA	Cooper C.S.;		
RT	"The t(X;1)(p11.2;q21.2) translocation in papillary renal cell		
RT	carcinoma fuses a novel gene PRC3 to the TFE3 transcription factor		
RT	gene."		
RL	Hum. Mol. Genet. 5:1333-1338(1996).		
RN	[3]		
RP	REVISIONS.		
RA	Clark J.;		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE OF 208-743 FROM N.A.		
RX	MEDLINE: 90249724.		
RA	Beckmann H., Su L.-K., Kadesch T.;		
RT	"TEF3: a helix-loop-helix protein that activates transcription		
RT	through the immunoglobulin enhancer muez motif."		
RL	Genes Dev. 4:167-179(1990).		
RN	[5]		
RP	SEQUENCE OF 266-353 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.		
RX	MEDLINE: 98054131.		
RA	Clark J., Lu Y.-J., Sihar S.K., Parker C., Gill S., Smedley D.,		
RA	Hamoudi R., Linehan W.M., Shipley J., Cooper C.S.;		
RT	"Fusion of splicing factor genes PSF and Nono (p54nbp) to the TFE3		
RT	gene in papillary renal cell carcinoma."		
RL	Oncogene 15:2233-2239(1997).		
CC	-1- FUNCTION: POSITIVE-ACTING TRANSCRIPTION FACTOR THAT BINDS TO THE		
CC	IMMUNOGLOBULIN ENHANCER MUEZ MOTIF. IT BINDS ALSO VERY WELL TO A		
CC	CC/MUTE SITE. BINDING OF TFE3 TO DNA INDUCES DNA BINDING.		
CC	-1- SUBUNIT: EFFICIENT DNA BINDING. REQUIRES DIMERIZATION WITH ANOTHER		
CC	CC BLH PROTEIN.		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.		
CC	-1- TISSUE SPECIFICITY: UBIQUITOUS IN FETAL AND ADULT TISSUES.		
CC	-1- DISEASE: INVOLVED IN PAPILLARY RENAL CELL CARCINOMA (PRCC) BY		
CC	CHROMOSOMAL TRANSLOCATIONS T(X;1)(P11.2;Q21.2) WHICH INVOLVES TFE3		
CC	AND PRC3: T(X;1)(P11.2;P34) WHICH INVOLVES TFE3 AND PSF, AND		
CC	INV(X)(P11.2;Q12) THAT INVOLVES TFE3 AND NONO.		

```

CC      -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC      TRANSCRIPTION FACTORS. BHLH-TIP SUBFAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X97121; CAA68061.1; -
DR      EMBL: X97160; CAA68800.1; -
DR      EMBL: X97161; CAA65800.1; JOINED.
DR      EMBL: X97162; CAA65800.1; JOINED.
DR      EMBL: X96717; CAA65478.1; -
DR      EMBL: X51330; CAA35714.1; -
DR      PIR: A34596; A34596.
DR      PIR: S10379; S10379.
DR      HSSP: P22415; 1AN4.
DR      TRANSFAC: T00811; -
DR      MIM: 314310; -
DR      PFAM: PF00010; HLH. 1.
DR      PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW      Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW      Chromosomal translocation; Proto-oncogene.
FT      DOMAIN 260 271
FT      (POTENTIAL).
FT      STRONG TRANSCRIPTION ACTIVATION DOMAIN
FT      (POTENTIAL).
FT      BASIC DOMAIN.
FT      HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT      LEUCINE-ZIPPER (POTENTIAL).
FT      PRO/ARG-RICH.
FT      BREAKPOINT FOR TRANSLLOCATION TO FORM
FT      PROC-TREE3 ONCOGENE.
FT      BREAKPOINT FOR TRANSLLOCATION TO FORM
FT      PSF-TREE3 ONCOGENE.
FT      BREAKPOINT FOR TRANSLLOCATION TO FORM
FT      SITE 295 296
FT      SITE 295 296
FT      SITE 295 296
FT      SITE 295 296
FT      CONFLICT 208 211
FT      CONFLICT 219 219
FT      CONFLICT 222 222
FT      CONFLICT 443 443
FT      CONFLICT 455 455
FT      CONFLICT 475 475
FT      CONFLICT 593 593
FT      CONFLICT 557 725
FT      CONFLICT 726 743
FT      MISSING (IN REF. 3).
FT      ESSNGGDSPGGLSSAPSP -> PAVSKASSRRSSFSMEES
FT      (IN REF. 3).
SQ      SEQUENCE 743 AA; 80007 MW; 1D82B94553543949 CRC64;
Query Match 1.8%; Score 143; DB 1; Length 743;
Best Local Similarity 33.8%; Pred. No. 1.82e-05;
Matches 26; Conservative 23; Mismatches 23; Indels 5; Gaps 4;
Db 346 OKKDHNHLEIERRRRINDRILEKLTGLTLLPKSSDPMRNKNGILKASVDYIRKLOKEOOR 405
Oy 355 EKRTAHNLEIKKYRCSINDRIOLQKVLV--CGD-EAKLSKSTLRALHIEFVEHENV 411
Db 406 SKD-LESRORSLEOANR 421
Oy 412 LKHVHEOMRKTL-QNNR 427
RESULT 9
ID N075-SOYBN STANDARD. PRT: 309 AA.
AC P08297;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EARLY MODULIN 75 PRECURSOR (N-75) (NGM-75).
GN ENOD2A AND ENOD2B.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

```





```
CC TRANSCRIPTION FACTORS. "MYOGENIC FACTORS" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M69052; AAA30009.1; ALT_INIT.
DR PIR; A41123; A41123.
DR HSSP; P10085; IMDY.
DR TRANSFAC; T00926; -.
DR PFAM; PF01586; Basic; 1.
DR PFAM; PF00010; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 112 124 BASIC DOMAIN.
FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 260 AA; 28682 MW; 8CC5FA9D2E6DAFA4 CRC64;

Query Match 1.6%; Score 128; DB 1; Length 260;
Best Local Similarity 33.8%; Pred. No. 2.96e-03;
Matches 23; Conservative 18; Mismatches 23; Indels 4; Gaps 4;

Db 112 DKRKATLRERRRLR-KVNEAFELKRRHCANPQRLPKVELIRNAIEYIEKLERLQVE 170
QY 355 ERTAHNLIEKK-YKCSINDRIQQLKVLGGD-EAKLSATLRARAIHIEVEHENQVL 412
Db 171 KANGSEM 178
QY 413 KHAYE-QM 419

RESULT 14
ID USF1_HUMAN STANDARD; PRT; 310 AA.
AC P22415;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UPSTREAM STIMULATORY FACTOR 1 (MAJOR LATE TRANSCRIPTION FACTOR 1).
GN USF1 OR USF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RX MEDLINE; 91065519.
RA Gregor P.D., Savadogo M., Roeder R.G.;
RT "The adenovirus major late transcription factor USF is a member of
RT the helix-loop-helix group of regulatory proteins and binds to DNA as
RT a dimer."
RL Genes Dev. 4:1730-1740(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 197-260.
RX MEDLINE; 94139661.
RA Ferre-D'Amaré A.R., Pogononec P., Roeder R.G., Burley S.K.;
RT "Structure and function of the b/HLH/Z domain of USF."
RL EMBO J. 13:180-189(1994).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO A SYMMETRICAL DNA
CC SEQUENCE (E-BOXES) (5'-CACGTG-3') THAT IS FOUND IN A VARIETY OF
CC VIRAL AND CELLULAR PROMOTERS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER
CC (USF1/USF2).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55666; CA39201.1; -.
DR PIR; S13525; S13525.
DR PDB; 1AN4; 17-SEP-97.
DR TRANSFAC; T00874; -.
DR MIM; 191523; -.
DR PFAM; PF00010; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; 3D-structure.
FT DNA_BIND 200 212 BASIC DOMAIN.
FT DOMAIN 213 255 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 271 292 LEUCINE-ZIPPER (POTENTIAL).
SQ SEQUENCE 310 AA; 33538 MW; BFDA91519BAB80AE CRC64;

Query Match 1.6%; Score 126; DB 1; Length 310;
Best Local Similarity 31.4%; Pred. No. 5.68e-03;
Matches 27; Conservative 25; Mismatches 27; Indels 7; Gaps 6;

Db 193 RTTDEKRRAGHNEVERRRRKINNIWVLSKIIPDCSMESTKSGSKGGLSKACDYI 252
QY 350 RRPKTE-RRTAHNLIEKKYKCSINDRIQQL-KVLL-CG-DEAK-LSKSATLRARAIHIE 403
Db 253 ELROSNRRLSEELQDLOLDNDVLL 278
QY 404 EVEHENQVLKHVQMRK-TLQNNRL 428

RESULT 15
ID USF1_RABIT STANDARD; PRT; 310 AA.
AC 002818; 002819;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UPSTREAM STIMULATORY FACTOR 1 (MAJOR LATE TRANSCRIPTION FACTOR 1).
GN USF1.
OS Eucytolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NEW ZEALAND WHITE; TISSUE=LUNG;
RX MEDLINE; 97435316.
RA Gao E., Wang Y., Alcorn J.L., Mendelson C.R.;
RT "The basic helix-loop-helix-zipper transcription factor USF1
RT regulates expression of the surfactant protein-A gene."
RL J. Biol. Chem. 272:23398-23406(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO A SYMMETRICAL DNA
CC SEQUENCE (E-BOXES) (5'-CACGTG-3') THAT IS FOUND IN A VARIETY OF
CC VIRAL AND CELLULAR PROMOTERS. REGULATES THE EXPRESSION OF THE
CC SURFACTANT PROTEIN-A (SP-A) GENE.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS AN HOMODIMER OR A HETERODIMER
CC (USF1/USF2).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: USF1A (SHOWN HERE) AND USF1B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003894; AAC48764.1; -.
DR EMBL; AF003895; AAC48765.1; -.

```

DR PFAM: PF00010: HLH; 1.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation;  
KW Alternative splicing.  
FT DNA\_BIND 200 212 BASIC DOMAIN.  
FT DOMAIN 213 255 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 271 292 LEUCINE-ZIPPER (POTENTIAL).  
FT VARSPLIC 131 158 MISSING (IN ISOFORM USF1B).  
SO SEQUENCE 310 AA; 33510 MW; 46EE3EF6FE08E0E CRC64.

Query Match 1.6%; Score 126; DB 1; Length 310;  
Best Local Similarity 31.4%; Pred. No. 5.68e-03;  
Matches 27; Conservative 25; Mismatches 27; Indels 7; Gaps 6;

Db 193 RTTDEKRRRAQHNEVERRRRDKINMWIYQLSKTIIPDCSMESTKSGQSGILSKACDYIQ 252  
QY 350 RRPKTE-RTAHNLIEKKYRCISINDRIQL-KVLL-CG-DEAK--LSKSATLRAIEHIE 403  
Db 253 ELRQSNHRLSEELQGLDQLDNDVYL 278  
QY 404 EVEHENQVYLKHHVQMRK-TLQNNRL 428

Search completed: Sun Sep 3 12:27:43 2000  
Job time : 62 secs.

THIS PAGE BLANK (USPTO)

RH 426 6466

RP 501 17